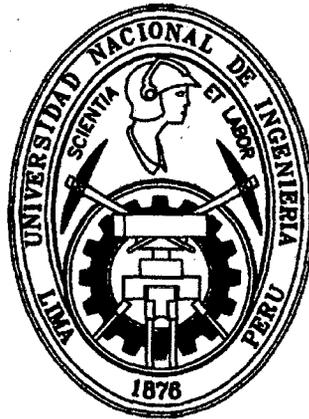


UNIVERSIDAD NACIONAL DE INGENIERÍA
FACULTAD DE INGENIERÍA INDUSTRIAL Y DE SISTEMAS
SECCIÓN DE POSGRADO



**MODELO DE SISTEMA DE SOPORTE PARA LA TOMA DE
DECISIONES PARA LA PUBLICACIÓN DE INFORMACIÓN EN
LA PÁGINA WEB DE UNA ENTIDAD PÚBLICA**

TESIS
PARA OPTAR POR EL GRADO ACADÉMICO DE
MAESTRO EN CIENCIAS CON MENCIÓN EN
INGENIERÍA DE SISTEMAS

ING. RONALD AUGUSTO CUELA ALVAREZ

ASESOR
DR. EMILIO ALBERTO UN JAN LIAU HING

LIMA – PERÚ

2012

Digitalizado por:

Consortio Digital del
Conocimiento MebLatam,
Hemisferio y Dalse

DEDICATORIA

A mis padres María Ceferina Alvarez Vega y Silverio Cuela Paiva.

A mi esposa María Isabel Raymundo Taipe.

A mi hijo Jair y a Joaquín que está en camino.

A mis familiares.

AGRADECIMIENTOS

Agradezco los consejos acertados de mi asesor Dr. Alberto Un Jan, los comentarios de mis jurados Mg. Celedonio Méndez y Mg. Paul Tocto, y los aportes de mis amigos de la maestría que me ayudaron en el desarrollo de esta tesis.

Agradezco a mi familia, sin su apoyo no hubiese podido culminar esta investigación.

ÍNDICE

ÍNDICE	iv
DESCRIPTORES TEMÁTICOS	xvi
RESUMEN	xvii
ABSTRACT.....	xix
INTRODUCCIÓN.....	1
CAPÍTULO 1 PROTOCOLO DE INVESTIGACIÓN.....	4
1.1 DIAGNÓSTICO Y EVALUACIÓN DEL PROBLEMA.....	4
1.2 DEFINICIÓN DEL PROBLEMA	7
1.2.1 PROBLEMA GENERAL	7
1.2.2 PROBLEMAS ESPECÍFICOS.....	7
1.3 IMPORTANCIA Y JUSTIFICACIÓN DE LA INVESTIGACIÓN.....	7
1.3.1 JUSTIFICACIÓN TEÓRICA	8
1.3.2 JUSTIFICACIÓN METODOLÓGICA	9
1.3.3 JUSTIFICACIÓN PRÁCTICA	9
1.4 HIPÓTESIS	10
1.4.1 HIPÓTESIS GENERAL.....	10

1.4.2	HIPÓTESIS ESPECÍFICAS	10
1.5	IDENTIFICACIÓN DE VARIABLES	10
1.5.1	VARIABLE INDEPENDIENTE.....	10
1.5.2	VARIABLE DEPENDIENTE	10
1.6	MATRIZ DE CONSISTENCIA.....	10
1.7	OBJETIVOS	11
1.7.1	OBJETIVOS GENERAL.....	11
1.7.2	OBJETIVOS ESPECÍFICOS	11
1.8	ALCANCES DE LA TESIS.....	12
CAPÍTULO 2 MARCO TEÓRICO.....		13
2.1	SISTEMA DE SOPORTE A DECISIONES (SSD).....	13
2.1.1	DEFINICIÓN DE UN SSD.....	13
2.1.2	PROCESO DE TOMA DE DECISIONES	14
2.1.3	CARACTERÍSTICAS DE UN SSD	15
2.1.4	SISTEMAS DE INFORMACIÓN Y SSD	16
2.1.5	COMPONENTES DE UN SSD.....	17
2.1.6	DISEÑO DE UN SSD.....	18
2.2	DATA WAREHOUSE (DW)	19
2.2.1	DEFINICIÓN DE DW	19
2.2.1.1	ENFOQUE DESCENDENTE (<i>TOP-DOWN</i>).....	19
2.2.1.2	ENFOQUE ASCENDENTE (<i>BOTTOM-UP</i>).....	20

2.2.2	CARACTERÍSTICAS DE UN DW.....	21
2.2.3	DIFERENCIAS CON LA BASE DE DATOS CONVENCIONAL	21
2.2.4	COMPONENTES DE UN DW	23
2.3	DATA MART.....	26
2.3.1	DEFINICIÓN DE DATA MART	26
2.3.2	CARACTERÍSTICAS DE UN DATA MART	27
2.3.3	TIPOS DE DATA MART.....	27
2.3.4	TOPOLOGÍA DE UN DATA MART.....	27
2.3.4.1	TOPOLOGÍA DE UNA CAPA	28
2.3.4.2	TOPOLOGÍA DE DOS CAPAS.....	28
2.3.4.3	TOPOLOGÍA DE TRES CAPAS.....	28
2.3.5	DESARROLLO DE DATA MART	29
2.4	ANÁLISIS DE USO DE UN SITIO WEB	29
2.4.1	RECOLECCIÓN DE DATOS PARA ANALIZAR EL USO DE UN SITIO WEB	30
2.4.1.1	USO DE REGISTROS DEL SERVIDOR WEB	30
2.4.1.2	USO DE SERVIDORES RECOLECTORES EXTERNOS (WEB BEACONS)	30
2.4.1.3	USO DE ETIQUETAS JAVASCRIPT.....	30
2.4.1.4	USO DE DETECTORES DE PAQUETES	30
2.4.2	REGISTROS DE UN SERVIDOR WEB	30

2.4.3	TIPOS DE DATOS PARA EL ANÁLISIS DE REGISTROS DE UN SITIO WEB	32
2.4.4	CONSIDERACIONES SOBRE EL ANÁLISIS DE UN SITIO WEB MEDIANTE REGISTROS DE SERVIDORES.....	33
2.4.5	BENEFICIOS DE UTILIZAR REGISTROS DE SERVIDORES PARA EL ANÁLISIS DE USO DE UN SITIO WEB	34
2.5	ENTIDADES PÚBLICAS	36
2.5.1	DEFINICIÓN	36
2.5.2	PORTAL DE INTERNET DE UNA ENTIDAD PÚBLICA	36
2.5.3	PORTAL DEL ESTADO PERUANO (PEP)	36
2.5.4	PORTAL DE SERVICIOS AL CIUDADANO Y EMPRESAS (PSCE) ..	37
2.5.5	TRANSPARENCIA Y ACCESO A LA INFORMACIÓN.....	38
2.5.6	PORTAL DE TRANSPARENCIA ESTÁNDAR	38
2.5.7	ACCESIBILIDAD DE LOS SITIOS WEB DE ENTIDADES PÚBLICAS	39
CAPÍTULO 3 ESTADO DEL ARTE.....		40
3.1	CLASIFICACIÓN DE LA INVESTIGACIÓN	40
3.2	ANTECEDENTES	41
3.2.1	NORMALIZACIÓN DE LOS REGISTROS DE UNA PÁGINA WEB....	41
3.2.2	UN DATA MART PARA EL ANÁLISIS DE HITS.....	42
3.2.3	ANÁLISIS DE USO DE UNA PÁGINAS WEB	43
3.2.4	ANÁLISIS DE SITIOS WEB SEGÚN USUARIO.....	45

3.2.5	MODELO PARA ANÁLISIS DE VISITAS Y PÁGINAS CON DOS ESTRELLAS	47
3.2.6	MODELO GENÉRICO PARA ANÁLISIS DE VISITAS Y PÁGINAS....	49
3.2.7	ESQUEMA DE ESTRELLA DETALLADO PARA LA NAVEGACIÓN..	49
3.3	BASE DEL MODELO SSD BUSCADO	51
3.4	APLICACIONES SIMILARES	51
3.4.1	AWSTATS.....	51
3.4.2	ANALOG	52
3.4.3	GOOGLE ANALYTICS.....	52
3.5	INDICADORES UTILIZADOS PARA EL ANÁLISIS DE UN SITIO WEB.....	52
CAPÍTULO 4	MODELO SSD-IWEP	53
4.1	ANÁLISIS INICIAL.....	53
4.1.1	OBJETIVO	53
4.1.2	LÍMITES DEL MODELO.....	54
4.2	ANÁLISIS SITUACIONAL.....	55
4.2.1	MODELO DE FUENTE DE DATOS	55
4.2.1.1	DATOS DE LA ESTRUCTURA DEL SITIO WEB	55
4.2.1.2	DATOS PROVENIENTES DE LA IP.....	55
4.2.1.3	DATOS PROVENIENTES DEL AGENTE DE NAVEGACIÓN	56
4.2.1.4	DATOS PROVENIENTES DE LOS REGISTROS DEL SERVIDOR WEB.....	57

4.2.2	ANÁLISIS DE REQUERIMIENTOS.....	58
4.2.3	DIAGRAMA GENERAL DEL SSD-IWEP.....	59
4.3	DISEÑO DEL SISTEMA.....	60
4.3.1	COMPONENTE DE DATOS.....	60
4.3.1.1	FLUJO DE DATOS.....	60
4.3.1.2	DISEÑO DE LA BASE DE DATOS.....	61
4.3.2	DISEÑO DE LOS MODELOS DE ANÁLISIS.....	61
4.3.2.1	MODELO PARA EL ANÁLISIS DE VISITANTES.....	61
4.3.2.2	MODELO PARA EL ANÁLISIS DE VISITAS.....	65
4.3.2.3	MODELO PARA ANÁLISIS DE PÁGINAS.....	69
4.3.3	INTERFACES DE USUARIO.....	72
4.3.3.1	REPORTES GEOGRÁFICOS.....	73
4.3.3.2	REPORTES DE VISITANTES INSTITUCIONALES.....	74
4.3.3.3	REPORTES DE CONTENIDO.....	75
4.3.3.4	REPORTES DE PÁGINAS DE ENTRADA.....	76
4.3.3.5	REPORTES DE PÁGINAS DE SALIDA.....	76
4.3.3.6	REPORTES DE REFERENCIAS.....	77
4.3.3.7	REPORTES DE SISTEMAS OPERATIVOS.....	78
4.3.3.8	REPORTES DE NAVEGADORES.....	79
4.3.3.9	REPORTES DE ACTIVIDAD EN EL SITIO WEB.....	79

4.3.3.10 OTROS REPORTES	80
4.3.4 CONSIDERACIONES EN LA ADMINISTRACIÓN Y MANTENIMIENTO DEL SISTEMA.....	81
CAPÍTULO 5 EVALUACIÓN DEL SSD-IWEP	83
5.1 DESCRIPCIÓN DEL AMBIENTE DEL CASO DE ESTUDIO	83
5.2 DISEÑO DE LA VALIDACIÓN DE LOS REPORTES.....	83
5.3 RESULTADOS DE APROBACIÓN DE REPORTES	84
5.3.1 PROMEDIO DE APROBACIÓN DE UN REPORTE	84
5.3.2 GRADO DE APROBACIÓN DE UN REPORTE	85
5.3.3 PROMEDIO DE APROBACIÓN DEL MODELO	86
5.3.4 GRADO DE APROBACIÓN DEL MODELO	86
5.3.5 PROMEDIO DE APROBACIÓN DEL REPORTE MÁS VALIOSO	86
5.3.6 GRADO DE APROBACIÓN DEL REPORTE MÁS VALIOSO	86
5.4 CARACTERÍSTICAS DEL MODELO.....	86
5.4.1 NO INTRUSIVO	86
5.4.2 SEGURO	87
5.4.3 MANTIENE LA PRIVACIDAD	87
5.4.4 DESEMPEÑO	87
5.4.5 ESCALABILIDAD	87
5.4.6 TRANSPARENTE EN DISEÑO.....	87
5.4.7 INDEPENDIENTE DE LA HERRAMIENTA	88

5.4.8 EXTENSIBLE.....	88
5.4.9 EXACTO.....	88
5.4.10 ACCESIBLE.....	88
5.4.11 FILTRA ROBOTS.....	88
5.4.12 MANTIENE LA PROPIEDAD.....	88
5.4.13 IDENTIFICA TODAS LAS TRANSACCIONES.....	88
5.5 APORTES DEL MODELO.....	89
CONCLUSIONES Y RECOMENDACIONES.....	90
1. CONCLUSIONES.....	90
2. RECOMENDACIONES.....	91
GLOSARIO DE TÉRMINOS.....	92
BIBLIOGRAFÍA.....	97
ANEXOS.....	103
1. ENCUESTA ENVIADA A LOS RESPONSABLES DE CONTENIDO DE SITIOS WEB DE ENTIDADES PÚBLICAS.....	103
2. PANTALLAS DE LA ENCUESTA ENVIADA A LOS RESPONSABLES DE SITIOS WEB DE ENTIDADES PÚBLICAS.....	105
2.1. PANTALLA DEL REPORTE REGIONAL (GEOGRÁFICO).....	105
2.2. PANTALLA DEL REPORTE DE VISITANTES INSTITUCIONALES.....	105
2.3. PANTALLA DEL REPORTE DE CONTENIDO.....	106
2.4. PANTALLA DEL REPORTE DE PÁGINAS DE ENTRADA.....	106
2.5. PANTALLA DEL REPORTE DE PÁGINAS DE SALIDA.....	107

2.6.	PANTALLA DEL REPORTE DE REFERENCIAS.....	107
2.7.	PANTALLA DEL REPORTE DE SISTEMAS OPERATIVOS	108
2.8.	PANTALLA DEL REPORTE DE NAVEGADORES.....	108
2.9.	PANTALLA DEL REPORTE DE ACTIVIDAD POR HORAS	109
3.	RESPUESTAS DE LOS ADMINISTRADORES DE CONTENIDO DE SITIOS WEB DE ENTIDADES PÚBLICAS	110

ÍNDICE DE FIGURAS

Figura 1:	Suscripciones de acceso a Internet fijo.....	5
Figura 2:	Publicaciones sobre "Sistemas de soporte a decisiones" en el ámbito académico*.....	8
Figura 3:	Modelo del proceso de toma de decisiones de Simon	14
Figura 4:	Modelo del proceso de toma de decisiones de Slade	15
Figura 5:	Productos de sistemas de información	16
Figura 6:	Componentes de un SSD	17
Figura 7:	Los componentes del Sistema Data Warehouse	23
Figura 8:	Elementos básicos de un Data Warehouse	24
Figura 9:	Estructura de un Data Warehouse.....	25
Figura 10:	Base de datos normalizada para los registros de la página web.....	42
Figura 11:	Tabla de hechos	43
Figura 12:	Dimensiones del modelo de datos de seguimiento de clics	44

Figura 13: Data Mart para monitorear el servidor web	46
Figura 14: Data Mart para los productos.....	47
Figura 15: Data Mart para análisis de páginas.....	48
Figura 16: Data Mart para análisis de sesiones	48
Figura 17: Un modelo genérico de estrella para los datos de la página web.....	49
Figura 18: Un modelo de estrella de Data Mart orientado a páginas web	50
Figura 19: Tablas para la estructura del sitio	55
Figura 20: Tablas asociadas a la IP.....	56
Figura 21: Tablas asociadas al agente de navegación	57
Figura 22: Tablas provenientes de los registros del servidor web	58
Figura 23: Diagrama general del modelo SSD-IWEP.....	59
Figura 24: Flujo de datos	60
Figura 25: Estructura completa de la Base de Datos	61
Figura 26: Modelo para análisis de visitantes	62
Figura 27: Modelo para análisis de visitas	66
Figura 28: Modelo para análisis de páginas.....	69
Figura 29: Partes de los reportes del modelo SSD-IWEP	73
Figura 30: Diseño de reportes geográficos	74
Figura 31: Diseño de reportes de visitantes institucionales.....	75
Figura 32: Diseño de reportes de contenido	75

Figura 33: Diseño de reportes de páginas de entrada	76
Figura 34: Diseño de reportes de páginas de salida	77
Figura 35: Diseño de reportes de referencias	78
Figura 36: Diseño de reportes de sistemas operativos	78
Figura 37: Diseño de reportes de navegadores	79
Figura 38: Diseño de reportes de actividad en el sitio web	80
Figura 39: Módulos de gestión de datos	81
Figura 40: Módulos de gestión de modelos	82
Figura 41: Módulos de gestión de interfaces	82

ÍNDICE DE TABLAS

Tabla 1: Matriz de consistencia	11
Tabla 2: Resumen del marco expandido de Power para los SSD	18
Tabla 3: Diferencias de una base de datos convencional y un Data Warehouse	22
Tabla 4: Topologías de un Data Mart.....	28
Tabla 5: Clasificación de la Tesis	40
Tabla 6: Valores para el análisis de visitantes	62
Tabla 7: Dimensiones para el análisis de visitantes.....	63
Tabla 8: Valores para el análisis de visitas	66
Tabla 9: Dimensiones para el análisis de visitas.....	67

Tabla 10: Valores para el análisis de páginas.....	70
Tabla 11: Dimensiones para el análisis de páginas	70
Tabla 12: Promedio de aprobación de un reporte	84
Tabla 13: Grado de aprobación de un reporte	85

DESCRIPTORES TEMÁTICOS

Sistema de información

Toma de decisiones

Internet

Sitio web

Entidad pública

Estadísticas de uso

Data Mart

Data Warehouse

Base de datos

Modelo de análisis

RESUMEN

La legislación peruana fomenta el acceso a la información para los ciudadanos y considera que los Portales de Internet de las entidades públicas se han convertido en el medio de difusión de información más accesible y permite que más ciudadanos obtengan información especializada.

Los administradores de contenido de sitios web de entidades públicas requieren de herramientas que les permitan analizar la creciente actividad de sus sitios web y cumplir con las normativas vigentes en el país. Las herramientas convencionales van mostrando limitaciones para resolver las crecientes necesidades de información de los administradores de contenido de sitios web de entidades públicas.

Los sistemas de soporte a decisiones (SSD) han mostrado una gran capacidad para brindar apoyo a los tomadores de decisiones en distintas organizaciones. Los SSD ayudan a reconocer los problemas, generar alternativas de solución y a seleccionar la alternativa adecuada.

Muchos investigadores han estudiado el uso de Data Mart y Data Warehouse para identificar el perfil de los visitantes y la actividad que realizan en sitios web. Gran parte de estas investigaciones están orientadas a los sitios web con comercio electrónico.

Algunos investigadores han planteado el uso de dos estrellas para el análisis de uso de un sitio web, una para visitas y otra para páginas. Este esquema permite usar dimensiones más adecuadas para identificar mejor a los visitantes y para encontrar patrones de uso de las páginas del sitio web.

Esta investigación muestra la utilidad del modelo de sistema de soporte para la toma de decisiones para la publicación de información en la página web de una entidad pública (SSD-IWEP). El modelo SSD-IWEP desarrolla las tres partes de un Sistema de Soporte para la toma de decisiones (SSD): los datos, los modelos de análisis y las interfaces.

El modelo SSD-IWEP está orientado a dar soporte a los administradores de contenido de sitios web de entidades públicas y plantea el uso de 3 estrellas, una para el análisis de visitantes con 17 dimensiones y 6 hechos; otra para análisis de visitas con 15 dimensiones y 5 hechos; y la última para evaluar las páginas vistas (o descargas) con 14 dimensiones y 2 hechos.

El modelo SSD-IWEP permite conocer mejor la actividad de visitantes en una página web y ayuda a reconocer los patrones de descarga en los mismos sitios. El modelo separa la actividad de arañas de la web (robots) del tráfico que se considera humano (no robots).

La validación del modelo SSD-IWEP es realizada con una encuesta que permite medir la aceptación de los principales reportes del modelo y del modelo en sí. En todos los casos se muestra que hay una aceptación de todos los reportes y del modelo en general por parte de los responsables de los portales de Internet de entidades públicas.

El modelo permite mejorar las decisiones sobre la información a difundir en los sitios web, orientar publicaciones a distintos públicos, priorización de contenido a publicar o retirar, promoción del sitio en otras páginas web, estructura del sitio web, tomar acciones ante las visitas de robots y publicación de documentos.

El modelo presenta además las características de ser seguro, mantiene la privacidad de los visitantes, tiene desempeño y escalabilidad a medida, transparente en el diseño, independiente de la herramienta, extensible, exacto, accesible, filtra robots, mantiene la propiedad de los datos y los modelos, identifica todas las transacciones y es no intrusivo.

ABSTRACT

Peruvian law promotes access to information for citizens and believes that the Internet Portals of public entities have become the more accessible way to disseminate information and allows citizens to get more specialized material.

The content managers of public entities require tools to analyze the increasing activity of their websites and to enable them to comply with regulations in the country. Conventional tools are showing limitations to solve the growing information needs of the content managers of public websites.

The decision support systems (DSS) have shown a great capacity to support decision makers in different organizations. The DSS help to identify problems, generate alternative solutions and select the appropriate alternative.

Many researchers have studied the use of Data Mart and Data Warehouse to identify the profile of visitors and their activity on websites. Much of this research is oriented e-commerce sites.

Some researchers have suggested using two stars for the analysis of website usage, one for visitors and one for pages (or download). This scheme allows using appropriate dimensions to a better identifying of visitors and page downloads.

This research demonstrates the utility of the system model to support decision making for the publication of information on the website of a public entity (SSD-IWEP). The model SSD-IWEP develops the three parts of a support system for decision-making (SSD): data, models and interfaces.

The SSD-IWEP model is geared to supporting content managers of websites of public institutions and raises the use of 3 stars, one for the analysis of visitors with 17 dimensions and 6 facts, another for analysis of visits with 15 dimensions and 5 facts, and to evaluate the page views (or downloads) with 14 dimensions and 2 facts.

The model SSD-IWEP gain insight into the activity of visitors to a website and help identify the firing patterns in the same places. The model separates the activity of the web spider's traffic (robots) from human activity (not robots).

The validation of the SSD-IWEP model is performed by a survey to measure the acceptance of the main reports of the model and the model itself. In all cases it is shown that there is an acceptance of all reports and the overall model by those responsible for the portals of public entities.

The model allows better decisions to disseminate the information on websites, target publications to different audiences, prioritizing to post or remove content, promoting the site on other web pages, improve web site structure, take actions for robots and publishing documents.

The model also has the characteristics of being safe, maintains the privacy of visitors, has performance and scalability, transparent design, tool-independent, extensible, accurate, accessible, filtered robots, retains ownership of the data and models, identifies all transactions and is not intrusive.

INTRODUCCIÓN

Las entidades públicas difunden en sus portales de Internet publicaciones, estadísticas, eventos, contenido informativo, educativo, preventivo, entre otra información, que son de interés de los ciudadanos. Según la Presidencia del Consejo de Ministros considera que los portales electrónicos institucionales de las entidades públicas se han convertido en el medio de difusión de información más accesible, son instrumentos muy poderosos de promoción de la imagen institucional, permitiendo que más usuarios obtengan información especializada (Presidencia del Consejo de Ministros, 2006).

Las páginas web de las entidades públicas tienen una participación importante en el tráfico de Internet en el país¹ y éstas enfrentan la necesidad de identificar a los visitantes a sus páginas web, sus patrones de comportamiento y sus necesidades de información.

En los últimos años se observa un aumento sostenido de usuarios de Internet en el país. Los datos a junio de 2012 muestran que la penetración de Internet en hogares alcanzó el 18,9% frente a 3,7% en 2005. Por otro lado, el 37,0% de la población de 6 y más años de edad ha tenido acceso a Internet en el trimestre abril-junio de 2012, mientras que en 2007 el 31,1% en 2007 (INEI, 2012).

Los administradores de sitios web de entidades públicas requieren de herramientas más avanzadas que les permitan conocer la actividad de los visitantes a sus

¹ Las estadísticas de Alexa (<http://www.alexa.com/>) indican que la Sunat, el Banco de la Nación, el Ministerio de Educación se encuentran entre las páginas más visitadas en el país, ocupando los puestos 27, 61 y 68, respectivamente.

Portales para poder mejorar la toma de decisiones sobre la información a publicar, la estructura que deben utilizar y la forma de difundir información en sus sitios web.

Esta investigación plantea un Sistema de Soporte a Decisiones que ayude a los administradores a conocer el comportamiento de sus visitantes y robots, y con ello mejorar la toma de decisiones en la difusión de información.

La investigación está dividida en 5 capítulos

En el primer capítulo se realiza la presentación del protocolo de investigación, en éste se describe la definición del problema, la importancia de la investigación, la ejecución de la investigación y los límites de la misma.

En el segundo capítulo se describe el marco teórico de la investigación, en esta parte se muestra los sistemas y herramientas en las que se enmarca la investigación y comenta sobre las normas en torno a las entidades públicas en torno a Internet.

El tercer capítulo detalla las investigaciones relacionadas sobre el estudio de modelos utilizados para identificar los patrones de comportamiento de visitantes en un sitio web. El autor muestra además las limitaciones de estos modelos para enfrentar el problema de la investigación.

El cuarto capítulo desarrolla el modelo planteado por el autor para resolver el problema. En este se detalla el análisis y diseño del Modelo de sistema de soporte para la toma de decisiones para la publicación de información en la página web de una entidad pública (Modelo SSD-IWEP).

El quinto capítulo presenta los resultados de la aceptación de los principales reportes que el modelo permite obtener por parte de los administradores de contenido de sitios web de entidades públicas.

Posteriormente se presentan las conclusiones y recomendaciones obtenidas de la investigación.

Luego se describe la bibliografía utilizada en la investigación y finalmente se muestran los anexos.

CAPÍTULO 1

PROTOCOLO DE INVESTIGACIÓN

1.1 DIAGNÓSTICO Y EVALUACIÓN DEL PROBLEMA

Las leyes peruanas fomentan el acceso a la información de entidades públicas a los ciudadanos, la Constitución Política del Perú indica en su Artículo 2:

“Toda persona tiene derecho a:

[...]

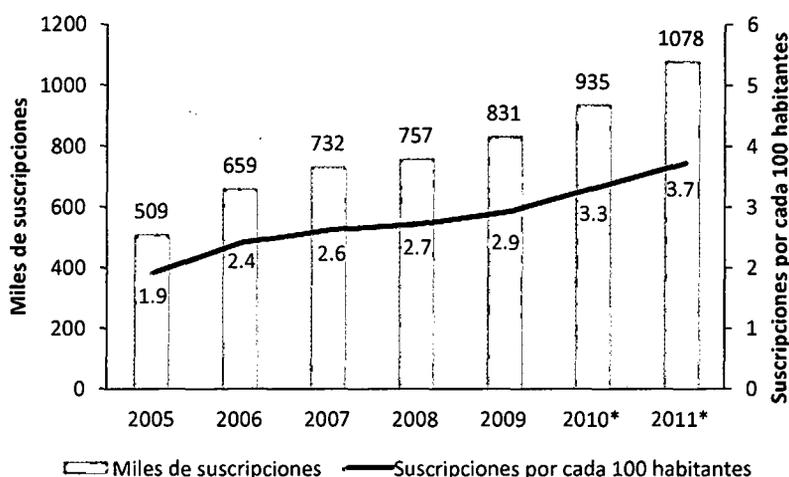
5. A solicitar sin expresión de causa la información que requiera y a recibirla de cualquier entidad pública, en el plazo legal, con el costo que suponga el pedido. [...]” (Congreso de la República del Perú, 1993).

Una de las principales funciones de las entidades públicas es la difusión de información. En los últimos años la difusión de información mediante sus páginas web se ha convertido en el medio de difusión de información más accesible, convirtiéndose en instrumentos de promoción de la imagen institucional, medio para aumentar las relaciones con los ciudadanos, fomentar los productos y servicios, y permite que más usuarios obtengan información especializada.

Para un responsable de una página web es importante determinar cuál es el origen y comportamiento de los visitantes de su página web y las características de las descargas de los documentos publicados por una entidad estatal. Esto les permite tener soporte para proponer mejoras en la página web y priorización en la creación de información. Por otro lado, también es importante filtrar los robots que visitan los sitios web y evaluar su comportamiento para poder optimizar la estructura y contenido de las páginas web.

En el país el acceso a Internet se ha ido incrementando en los últimos años. En 2005 el número de suscriptores con acceso a Internet fijo alcanzaba a 509 mil personas; en 2011, esta cifra estimada es de 1,078 millones. En la Figura 1 se muestra además que 3,7 personas de cada 100 tienen una suscripción a una línea fija en 2011 (OSIPTTEL, 2012).

Figura 1: Suscripciones de acceso a Internet fijo



Fuente: (OSIPTTEL, 2012).

La herramienta convencional de análisis de los registros del servidor web es el software de servidor AwStats. Esta herramienta permite visualizar los indicadores de una página web por meses, sin embargo los requerimientos de las estadísticas de uso se han hecho más complejos y se necesita responder a preguntas sobre el comportamiento de visitantes y descargas que involucran más de dos dimensiones (contenido, tiempo, lugar, navegador, duración de visitas), e incluso distintas agregaciones de estas dimensiones.

Otro aspecto es la identificación de robots (arañas de la web) que es necesario estudiarlos por dos razones: La primera es para poder filtrarlos de las estadísticas de uso de visitantes y descargas, y la segunda es para revisar el comportamiento de los robots en un sitio web.

Varios autores han propuesto el uso de modelos basados en Data Mart para el análisis de uso de páginas web de manera multidimensional y con distintas agregaciones. Sin embargo estos autores se han enfocado en páginas web de

instituciones privadas, cuyo objetivo principal es el análisis de su sección de comercio electrónico y de las compras realizadas en el sitio web. En una institución pública se requiere conocer la información que los ciudadanos están visitando con el propósito de fomentar dichas publicaciones, evaluar las secciones menos visitadas de un sitio web.

Por otro lado, un grupo importante de modelos plantea el uso de etiquetas JavaScript para el rastreo de información, esto lleva a un problema dado que muchos navegadores están incluyendo opciones para deshabilitar JavaScript debido a la vulnerabilidad que podrían causar o en otros casos pueden ser considerados software malicioso por algunos firewalls. Un aspecto más importante –que afecta al reconocimiento de descargas– es la existencia de enlaces externos que dirigen a los documentos (archivos PDF, XLS, ZIP, etc.) de un sitio web, dado que no es posible añadir etiquetas JavaScript a estos documentos, no es posible rastrearlos y se pierde gran parte de la información importante para la toma de decisiones.

Un aspecto adicional se debe al hecho de estudiar el comportamiento de los robots para evaluar si el contenido de un sitio web está siendo revisado o ignorado por ellos, que secciones son las más rastreadas y por último optimizar la accesibilidad del contenido desde buscadores.

Dada la sensibilidad de información que se utiliza en estos casos, se requiere además de un modelo no intrusivo, que evite interacción con los visitantes, y que a la vez permita recopilar información de su actividad en el sitio. Además se debe tener en cuenta los derechos de privacidad de visitantes.

Los administradores de sitios web de entidades públicas requieren de alternativas que ofrezcan la misma o mayor seguridad que las herramientas convencionales, un adecuado desempeño, que permitan enfrentar futuras necesidades y que permita la implementación en las plataformas disponibles, manteniendo las normas en las que están enmarcadas.

Ante ello, es necesario desarrollar un modelo que permita mejorar la toma de decisiones en la publicación de información en una entidad pública.

1.2 DEFINICIÓN DEL PROBLEMA

1.2.1 PROBLEMA GENERAL

PG: ¿Cómo mejorar la toma de decisiones en la publicación de información en una entidad pública?

1.2.2 PROBLEMAS ESPECÍFICOS

PE1: ¿Cómo mejorar la identificación de perfiles de visitantes de una página web en una entidad pública?

PE2: ¿Cómo mejorar la identificación de patrones de descarga de una página web en una entidad pública?

1.3 IMPORTANCIA Y JUSTIFICACIÓN DE LA INVESTIGACIÓN

Los administradores de contenido de páginas web de entidades públicas necesitan tener mayor conocimiento sobre el comportamiento de los visitantes y sus patrones de descarga de su página web para poder tomar decisiones sobre la información que difunden en sus portales.

Estos administradores enfrentan cada vez mayor información que analizar, debido a que el número de visitantes se va incrementando y con ello las descargas. Además los administradores de contenido cada vez van teniendo nuevos pedidos de información que requieren consultas ad hoc, consultas cruzadas, entre otros.

El estudio de comportamiento de robots se hace necesario debido a que una cantidad de ellos proviene de buscadores que hacen más accesible la información de un sitio web y por otra parte a veces generan distorsión en el número de visitas y descargas de un sitio web.

Los sistemas de soporte para la toma de decisiones para análisis de visitantes están difundidos en el ambiente empresarial, sin embargo el análisis de uso de páginas web de entidades públicas no está estudiado. Las entidades públicas presentan características distintas a las entidades privadas que hacen que sus

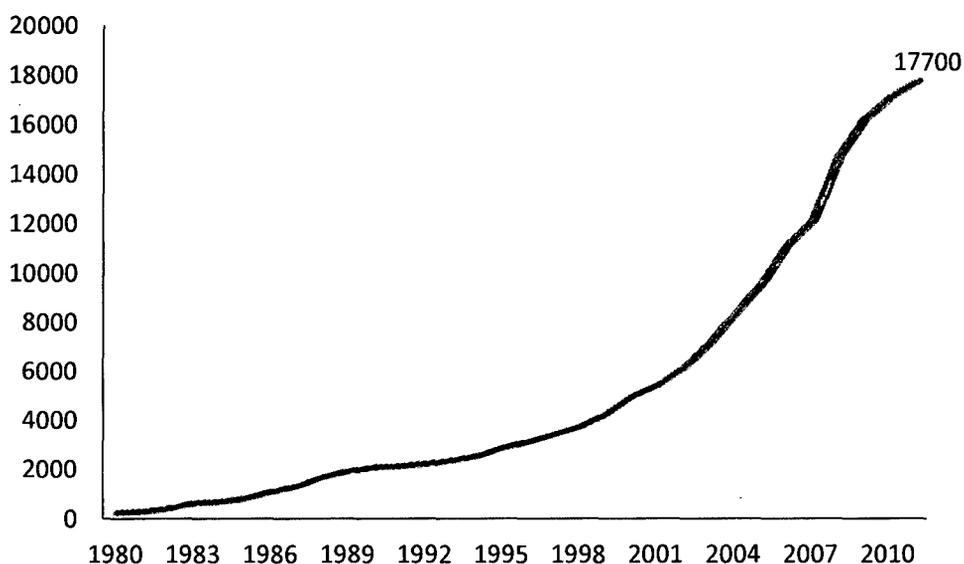
necesidades de información sean distintas y por lo tanto requieren de un modelo adecuado a sus necesidades.

Por otro lado, la investigación plantea algunas características que el modelo debe cumplir para aumentar la posibilidad de implementación de un modelo en distintas entidades del estado.

1.3.1 JUSTIFICACIÓN TEÓRICA

Los "Sistemas de Soporte a la Toma de Decisiones" son un tema de interés en el ámbito académico, en el Figura 2, se muestran las publicaciones que incorporan el término "*decision support systems*"², se observa que el este término cada año hay más publicaciones relacionadas y alcanza en 2011 la cifra de 17 700 resultados.

Figura 2: Publicaciones sobre "Sistemas de soporte a decisiones" en el ámbito académico*



* Tendencia de publicaciones que contienen "*decision support systems*" indexados por Google Scholar.

Fuente: (elaboración propia).

² El autor sólo utiliza el término en inglés debido a que la traducción al español para el término "*decisión support systems*" tienen una variedad no documentada (por ejemplo sistemas de soporte a decisiones, sistemas de soporte a toma de decisiones, sistemas de apoyo a tomas de decisiones, entre otros).

La investigación propone un modelo que extiende propuestas previas centradas en el análisis de datos de los registros de los servidores. Además, tiene en cuenta las necesidades de una entidad pública.

La investigación evalúa modelos previos orientados al comercio electrónico, y la dirige al análisis de las descargas de publicaciones de una entidad pública y sobre sus visitantes y robots.

Para la evaluación de la mejora en la toma de decisiones en una entidad pública, el autor plantea una encuesta a los administradores de contenido de páginas web de entidades públicas.

1.3.2 JUSTIFICACIÓN METODOLÓGICA

El comportamiento de los visitantes de una página web de una entidad pública tiene sus propias características frente al comportamiento de una empresa comercial. La tesis evaluará los distintos modelos planteados previamente para encontrar el modelo adecuado que permita analizar el comportamiento de este tipo de usuarios.

Las investigaciones previas no incorporan la identificación previa de las vistas de robots (arañas de la web), ni la parte de control de accesos que es muy importante en entidades que buscan evaluar la difusión de su información.

Las entidades públicas requieren tratamiento especial de la información debido a que tienen objetivos y marcos distintos a las entidades privadas. Esto requiere el planteamiento de un modelo que incluya los requerimientos generales de estas instituciones.

1.3.3 JUSTIFICACIÓN PRÁCTICA

El modelo resultante de la investigación brinda soporte a las decisiones de modificación de la página web y consecuentemente la priorización en la generación de la información.

El modelo permite evaluar el comportamiento de los robots y ello permitirá tomar mejores decisiones sobre la optimización del sitio, creación de contenido accesible para los buscadores y evaluación de otros tipos de robots.

El modelo SSD-IWEP es evaluado por los administradores de contenido de sitios web de entidades públicas.

1.4 HIPÓTESIS

1.4.1 HIPÓTESIS GENERAL

HG: El modelo SSD-IWEP mejorará la toma de decisiones en la publicación de información en una entidad pública.

1.4.2 HIPÓTESIS ESPECÍFICAS

HE1: El modelo SSD-IWEP mejorará la identificación de perfiles de visitantes de una página web en una entidad pública.

HE2: El modelo SSD-IWEP mejorará la identificación de patrones de descarga de una página web en una entidad pública.

1.5 IDENTIFICACIÓN DE VARIABLES

1.5.1 VARIABLE INDEPENDIENTE

El modelo SSD-IWEP.

1.5.2 VARIABLE DEPENDIENTE

Toma de decisiones en la publicación de información en una entidad pública.

1.6 MATRIZ DE CONSISTENCIA

La matriz de consistencia se muestra en la Tabla 1 y resume el problema, el objetivo e hipótesis.

Tabla 1: Matriz de consistencia

Problema	Objetivo	Hipótesis
PG: ¿Cómo mejorar la toma de decisiones en la publicación de información en una entidad pública?	OG: Mejorar la toma de decisiones en la publicación de información en una entidad pública.	HG: El modelo SSD-IWEP mejorará la toma de decisiones en la publicación de información en una entidad pública.
PE1: ¿Cómo identificar los perfiles de visitantes de una página web en una entidad pública?	OE1: Identificar los perfiles de visitantes de una página web en una entidad pública.	HE1: El modelo SSD-IWEP identificará los perfiles de visitantes de una página web en una entidad pública.
PE2: ¿Cómo identificar los patrones de descarga de una página web en una entidad pública?	OE2: Identificar los patrones de descarga de una página web en una entidad pública.	HE2: El modelo SSD-IWEP identificará los patrones de descarga de una página web en una entidad pública.

Fuente: (Elaboración propia).

1.7 OBJETIVOS

1.7.1 OBJETIVOS GENERAL

OG: Mejorar la toma de decisiones en la publicación de información en una entidad pública.

1.7.2 OBJETIVOS ESPECÍFICOS

OE1: Mejorar la identificación de perfiles de visitantes de una página web en una entidad pública.

OE2: Mejorar la identificación de patrones de descarga de una página web en una entidad pública.

1.8 ALCANCES DE LA TESIS

La investigación abarcará las fases de análisis y diseño del modelo de SSD-IWEP. Esta tesis desarrolla los componentes de un Sistema de Soporte para la toma de Decisiones planteado por Sauter: Datos, modelos e interfaces (Sauter, 2010).

La investigación abarca las respuestas de los administradores de contenido de páginas web de entidades públicas listadas en la página del Portal del Estado Peruano (<http://www.peru.gob.pe>). Los resultados de la investigación toman en cuenta las respuestas de los administradores de contenido de páginas web. La investigación evalúa los reportes que se pueden obtener del modelo planteado.

Algunos aspectos no se publican de acuerdo a las normativas legales, de ética, de confidencialidad y de las normas de seguridad informática de la entidad pública, vigentes durante la investigación.

CAPÍTULO 2

MARCO TEÓRICO

2.1 SISTEMA DE SOPORTE A DECISIONES (SSD)

“Los sistemas computarizados de apoyo a las decisiones se convirtieron en práctica [usual] con el desarrollo de las minicomputadoras, sistemas de aprovechamiento por turno de operación, y la computación distribuida. La historia de la aplicación de estos sistemas comienza a mediados de la década de 1960.” (Power D. , 2008b, p. 121).

2.1.1 DEFINICIÓN DE UN SSD

La definición propuesta por (Power D.) y tomada por (Power, Burstein, & Sharda, 2011) indica que un SSD es un sistema informático interactivo o subsistema con el objeto de ayudar a los tomadores de decisiones a usar las tecnologías de comunicaciones, datos, documentos, conocimiento y/o modelos para identificar y resolver problemas, completar tareas del proceso de decisión, y tomar decisiones. Un SSD es un término general para cualquier aplicación informática que mejora la capacidad de una persona o grupo para tomar decisiones.

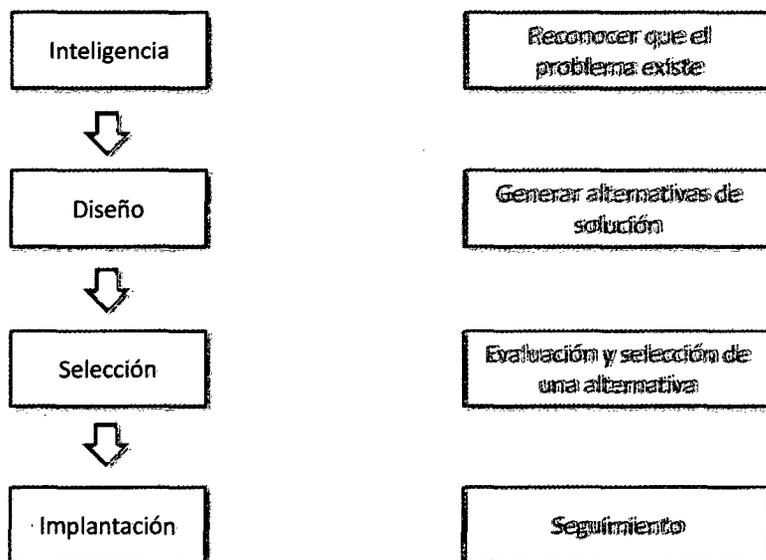
Los SSD tienen como finalidad apoyar la toma de decisiones mediante la generación y evaluación sistemática de alternativas o escenarios de decisión mediante el empleo de modelos y herramientas computacionales. Un SSD no soluciona problemas, ya que solo apoya el proceso de toma de decisiones. La responsabilidad de tomar una decisión, de adoptarla y de ponerla en práctica es de los administradores, no del SSD (Cohen & Asín, 2000, p. 188).

2.1.2 PROCESO DE TOMA DE DECISIONES

El modelo de Simon (Cohen & Asín, 2000) representado en la Figura 3 consta de cuatro fases:

1. Inteligencia: Se reconoce que existe un problema para el cual se debe tomar una decisión.
2. Diseño: Se generan las alternativas de solución para el problema identificado en la fase anterior.
3. Selección: Se evalúan las alternativas generadas en la fase anterior y se escoge la mejor.
4. Implantación: Se pone en marcha y se monitorea la alternativa seleccionada.

Figura 3: Modelo del proceso de toma de decisiones de Simon



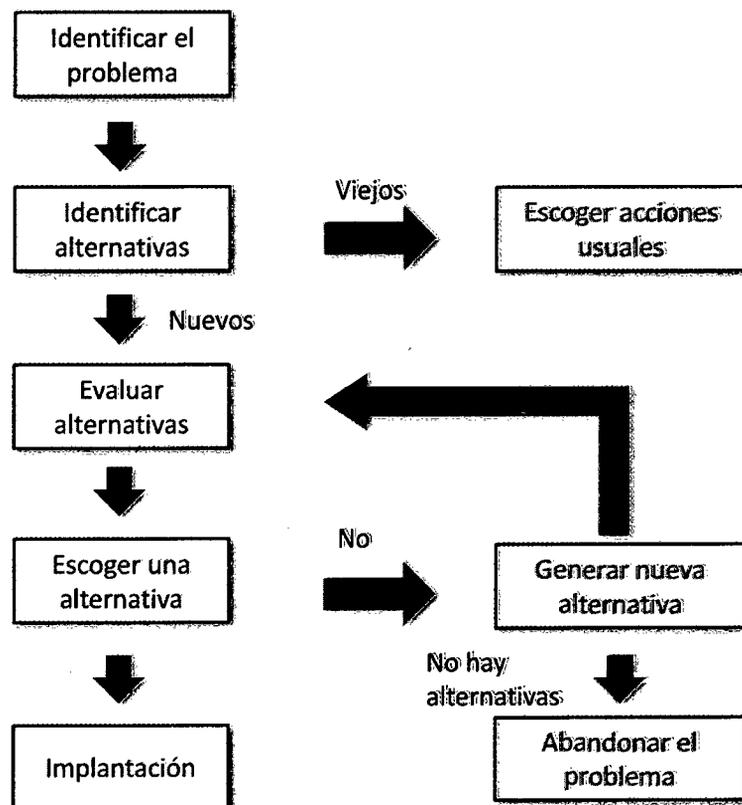
Fuente: (Cohen & Asín, 2000)

El modelo de Slade (Cohen & Asín, 2000) mostrado en la Figura 4 tiene las siguientes fases:

1. Identificación del problema.
2. Identificación de alternativas de solución. Aquí una distinción, si los problemas son viejos (anteriormente resueltos) se procede tomando las acciones usuales. Si el problema es nuevo pasa a la siguiente fase.

3. Evaluación de alternativas.
4. Elección de una alternativa de acuerdo a los requerimientos de la empresa.
En caso no se encuentre una alternativa adecuada se sigue hasta encontrar una o se decide que la solución no es factible y se abandona el problema. Si hay una alternativa factible se procede con el siguiente paso.
5. Implantación de la alternativa.

Figura 4: Modelo del proceso de toma de decisiones de Slade



Fuente: (Cohen & Asín, 2000)

Un Sistema de Soporte de Decisiones puede utilizarse como apoyo en las 3 primeras etapas del proceso de toma de decisiones de Simon (Figura 3) o en las cuatros primeras del modelo de Slade (Figura 4).

2.1.3 CARACTERÍSTICAS DE UN SSD

Las características necesarias para poder considerar a un Sistema como SSD (Cohen & Asín, 2000) son:

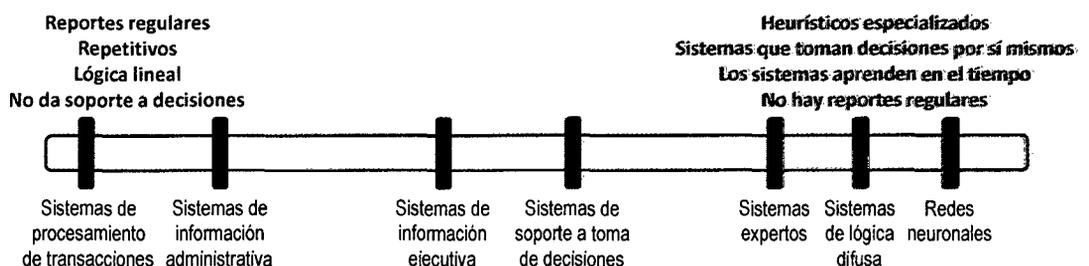
- Interactividad
- Tipo de decisiones
- Frecuencia de uso
- Variedad de usuarios
- Flexibilidad
- Desarrollo de modelos de decisión
- Interacción ambiental
- Comunicación interorganizacional
- Acceso a bases de datos
- Simplicidad

2.1.4 SISTEMAS DE INFORMACIÓN Y SSD

“Un sistema de información es un conjunto de elementos que interactúan entre sí con el fin de apoyar las actividades de una empresa o negocio. En un sentido amplio, un sistema de información no necesariamente incluye equipo electrónico (*hardware*). Sin embargo, en la práctica se utiliza como sinónimo de ‘sistema de información computarizado’.” (Cohen & Asín, 2000).

Un SSD es un Sistema de Información. La ubicación de los SSD dentro de los productos de sistemas de información con respecto a toma de decisiones se muestra en la Figura 5. Los SSD están en el intermedio de los SI que producen reportes repetitivos y los sistemas especializados que toman decisiones.

Figura 5: Productos de sistemas de información



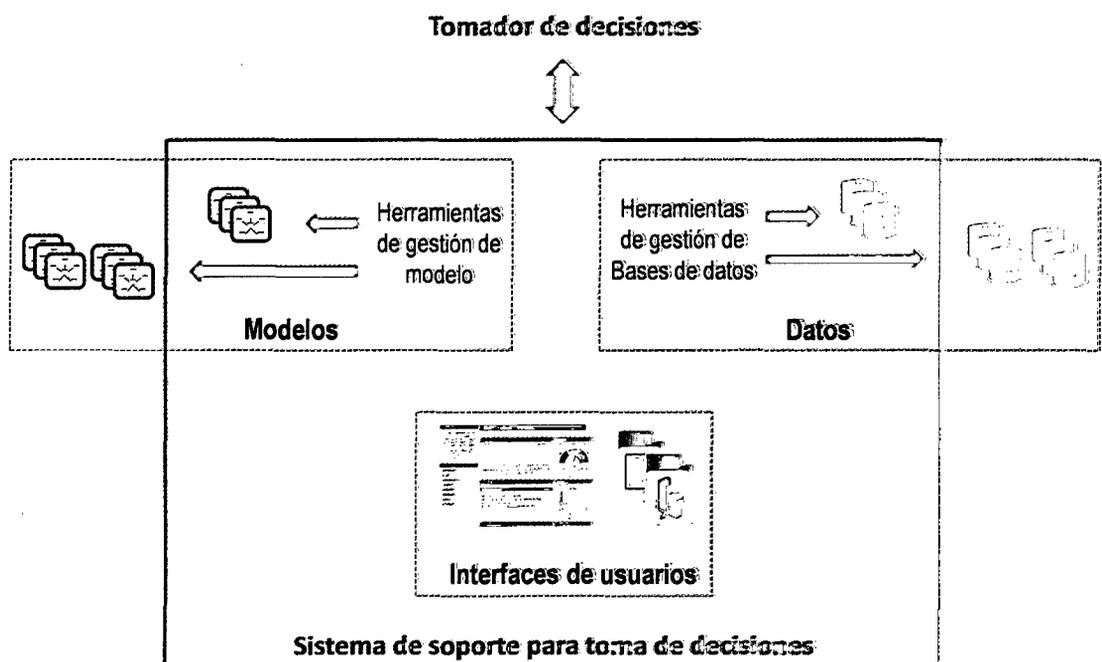
Fuente: (Sauter, 2010).

2.1.5 COMPONENTES DE UN SSD

Los componentes de un SSD se muestran en la Figura 6, estos son:

- El modelo:
- Los datos:
- Interface de usuario:

Figura 6: Componentes de un SSD



Fuente: (Sauter, 2010).

Un concepto propuesto por Sprague & Carlson y tomado por (Power D. , 2008a), incluye cuatro componentes:

1. La interfaz de usuario
2. La base de datos
3. Los modelos y herramientas analíticas
4. La arquitectura y red del SSD

Power (Power D. , 2008a) propone cinco componentes en un marco ampliado de cuatro dimensiones (una primaria y tres secundarias) enfocado básicamente en las

diferencias de los DSS cuando ya están contruidos (Tabla 2). La importancia de cada componente en la prestación de funcionalidades para el apoyo a la toma de decisiones es el principal factor diferenciador entre las distintas categorías de SSD.

Tabla 2: Resumen del marco expandido de Power para los SSD

Componente Dominante	Grupo usuario	Propósito	Tecnología utilizada
Comunicaciones SSD orientado a comunicaciones	Usuarios intra e inter organizacionales	Conducir una reunión, publicar en un boletín	Basado en web o en una red local (LAN)
Base de datos SSD orientado a datos	Administradores, funcionarios, usuarios intra e inter organizacionales	Consultas en un Data Warehouse, Análisis ad hoc	Servidor, LAN, Basado en web
Base de Documentos SSD orientado a documentos	Especialistas, administradores	Búsqueda de páginas web, encontrar documentos	Basado en web
Base de Conocimiento SSD orientado a conocimiento	Usuarios internos, clientes	Asistencia a los administradores, selección de productos	LAN o Basado en web
Modelo SSD orientado a Modelos	Administradores, funcionarios, clientes	Calendario del equipo, Análisis de decisión	Aplicación de escritorio o Basado en web

Fuente: (Power D. , 2008a, p. 233).

2.1.6 DISEÑO DE UN SSD

La metodología para el diseño de un SSD tiene las siguientes fases (Sauter, 2010):

En la fase de análisis inicial busca entender las necesidades de información y el entorno. Esto implica determinar las decisiones claves y las necesidades de información clave.

En la fase de análisis de situación, se debe identificar las necesidades remanentes. Para ello se requiere entender las características de la organización, de las tareas y las características del usuario.

En la fase de diseño del sistema, se requiere hacer un diseño lógico del sistema, la construcción y evaluación del sistema.

En la fase de implementación se debe incorporar la demostración, el entrenamiento y desarrollo del sistema.

2.2 DATA WAREHOUSE (DW)

2.2.1 DEFINICIÓN DE DW

Los modelos principales de Data Warehousing son los de Inmon y Kimball (Breslin, 2004).

El enfoque de Inmon se orienta a un desarrollo descendente (*top-down*) que adapta las herramientas de bases de datos relacionales tradicionales a las necesidades de un Data Warehouse corporativo. Desde este Data Warehouse corporativo se desarrolla las bases de datos departamentales que brindan soporte a las decisiones y cubren la mayor parte de las necesidades de información de la empresa.

El enfoque de Kimball, por otro lado, propone un enfoque ascendente (*bottom-up*) que utiliza el modelado dimensional, en lugar de construir una única base de datos corporativa. Kimball propone la creación de Data Mart para los procesos de negocio más importantes.

2.2.1.1 ENFOQUE DESCENDENTE (TOP-DOWN)

Un Data Warehouse se define formalmente en los siguientes términos (Inmon W. , 2002):

- **Orientado a temas:** Los datos en el Data Warehouse están organizados de tal forma que las entidades están relacionados con el mundo real.
- **Históricos** (variantes en el tiempo): El ambiente de tiempo es mucho mayor que en el ámbito operacional, la estructura incluye implícita o explícitamente un elemento de tiempo y una vez que los datos están grabados correctamente, no pueden ser actualizados.
- **No volátil:** Los datos en el Data Warehouse no puede ser actualizada o borrada. Los datos almacenados son estáticos, de solo lectura y retenidos para reportes futuros.
- **Integrado:** El Data Warehouse almacena los datos de la mayor parte o de todas las aplicaciones operacionales de la organización y los datos son consistentes.

Posteriormente añada que se trata de “Una colección de datos que da soporte a decisiones a nivel gerencial.” (Inmon, Strauss, & Neushloss, 2008).

Inmon indica “... los datos en un Data Warehouse puede ser usado para muchos propósitos diferentes, incluyendo ... esperar para requerimientos futuros que no son conocidos ahora.” (Inmon W. , 2002, p. 29).

2.2.1.2 ENFOQUE ASCENDENTE (*BOTTOM-UP*)

Kimball proponiendo una aproximación dimensional para la construcción de Data Warehouse y provee una definición de resumida “Un Data Warehouse es un sistema que extrae, limpia, ajusta, y envía una fuente de datos en un almacén de datos dimensional y luego soporta e implementa consultas y análisis con el propósito de toma de decisiones” (Kimball & Ross, 2002).

Los objetivos de un Data Warehouse son:

- **Entendible:** El DW debe hacer que la información de la organización sea fácilmente accesible.
- **Creíble:** El DW debe presentar la información de una organización de manera consistente.
- **Adaptativo:** El DW debe ser adaptativo y elástico al cambio.

- Seguro: El DW debe ser un bastión seguro para proteger el activo de información.
- El DW debe servir como base para mejorar la toma de decisiones.
- Los integrantes de la organización deben aceptar el DW para considerarlo exitoso.

2.2.2 CARACTERÍSTICAS DE UN DW

Las características de un Data Warehouse (Turban & Volonino, 2011) son:

1. Organización. Los datos son organizados por temas y sólo contienen información relevante para la toma de decisiones.
2. Consistencia. Los datos en un DW son codificados de una manera consistente.
3. Histórico. Los datos son almacenados por muchos años con el objetivo de identificar tendencias, hacer regresiones y realizar comparaciones en el tiempo.
4. No volátil. Una vez que los datos han entrado al DW, ya no pueden ser modificados.
5. Relacional. Usualmente un DW utiliza una estructura relacional.
6. Cliente/servidor. El DW usa una arquitectura cliente/servidor principalmente para proveer a los usuarios un acceso fácil a sus datos.
7. Basado en web. Hoy en día los DW son diseñados para proveer un ambiente informático eficiente para aplicaciones basadas en Web.
8. Integración. Los datos de distintas Fuentes son integrados en el DW, los Servicios Web son utilizados para dar soporte a la integración.
9. Tiempo real. A pesar de la mayoría de un DW no brinda respuesta en tiempo real, es posible elaborar algunas funcionalidades que si lo hagan.

2.2.3 DIFERENCIAS CON LA BASE DE DATOS CONVENCIONAL

Las bases de datos convencionales son de procesamiento de transacciones en línea (OLTP) en la que cada transacción debe registrarse rápidamente. Las bases de datos son volátiles porque los datos se están agregando constantemente, editada o actualizada.

A diferencia de las bases de datos, los Data Warehouse no son volátiles. Por lo tanto, los Data Warehouse están diseñados como procesamiento analítico en línea (OLAP) de sistemas, lo que significa que los datos pueden ser consultados y analizados de manera más eficiente que las bases de datos de aplicaciones OLTP.

La Tabla 3 muestra las principales diferencias de una base de datos convencional y un DW.

Tabla 3: Diferencias de una base de datos convencional y un Data Warehouse

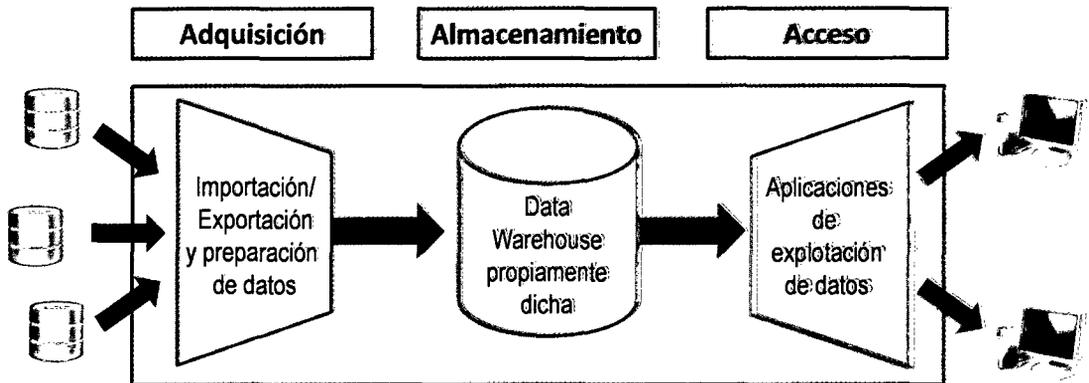
BASE DE DATOS	DATA WAREHOUSE
Predomina la actualización	Predomina la consulta
La actividad más importante es de tipo operativo (día a día)	La actividad más importante es el análisis y la decisión estratégica
Predomina el proceso puntual	Predomina el proceso masivo
Mayor importancia a la estabilidad	Mayor importancia al dinamismo
Datos en general desagregados	Datos en distintos niveles de detalle y agregación
Importancia del dato actual	Importancia del dato histórico
Importante del tiempo de respuesta de la transacción instantánea	Importancia de la respuesta masiva
Estructura relacional	Visión multidimensional
Usuarios de perfiles medios o bajos	Usuarios de perfiles altos
Explotación de la información relacionada con la operativa de cada aplicación	Explotación de toda la información interna y externa relacionada con el negocio

Fuente: (Inmon, Strauss, & Neushloss, 2008), (Inmon W., 2002), (Kimball & Ross, 2002).

2.2.4 COMPONENTES DE UN DW

Los principales componentes de un DW se muestran en la Figura 7 (Mattison, 2006).

Figura 7: Los componentes del Sistema Data Warehouse

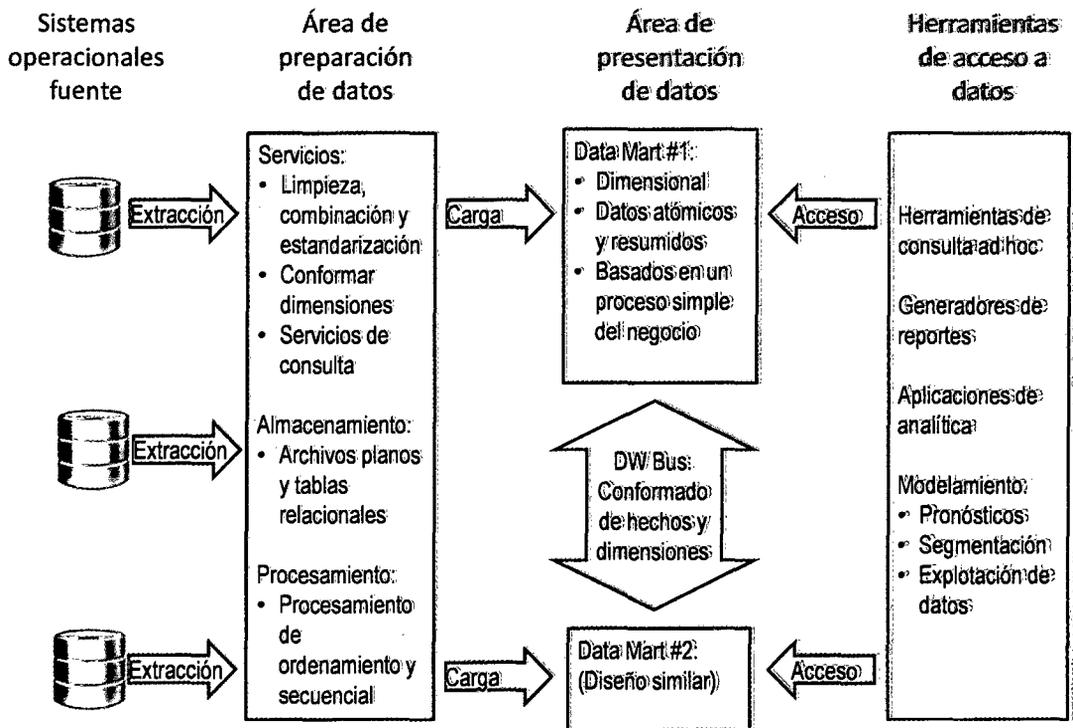


Fuente: (Mattison, 2006).

- El primer componente del sistema es el Data Warehouse (el almacén de datos propiamente dicho), Se trata de la base de datos física, que almacena una gran cantidad de información de una amplia variedad de fuentes. Los datos están organizados de tal forma que es fácil de usar y optimizada para búsquedas, y es actualizada frecuentemente desde sus fuentes.
- El segundo componente del sistema es el componente de adquisición de datos (importación/exportación y preparación). Esta parte contiene todas las herramientas, programas, interfaces de carga de sistemas transaccionales, preparación y carga al Data Warehouse, incluyendo exportación e importación del Data Warehouse cuando sea necesario.
- El tercer y más importante componente es el componente de acceso que incluye todas las aplicaciones que permiten utilizar la información del Data Warehouse.

Kimball presenta una versión más específica de los elementos de un Data Warehouse que se muestra en la Figura 8

Figura 8: Elementos básicos de un Data Warehouse



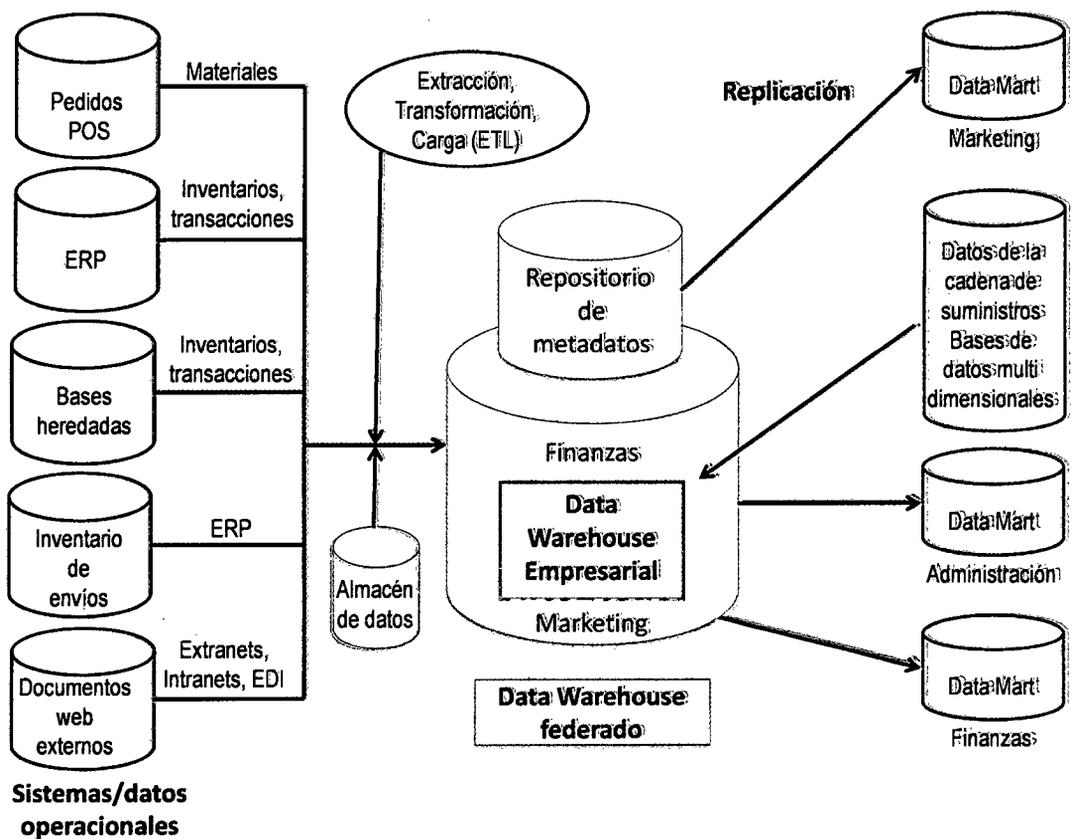
Fuente: (Kimball & Ross, 2002).

- Los sistemas operacionales tienen información de las transacciones del Negocio. Estos sistemas están fuera del alcance del DW. Son la fuente del DW.
- La zona de concentración de datos incluye un área de almacenamiento y un conjunto de procesos comúnmente denominados ETL (*Extract, Transformation, Load*) o zona de Extracción, Transformación y Carga de datos. Es posible usar bases de datos normalizadas en esta área, sin embargo para pasar a la siguiente zona generalmente se requiere que tengan una estructura dimensional.
- La Zona de presentación de datos es un lugar donde los datos son organizados, almacenados y preparados para consultas directas de los usuarios, generadores de reportes y otras aplicaciones analíticas. Usualmente esta área está compuesta de diversos Data Mart.
- Las Herramientas de acceso a los datos es el componente más importante. Estas incluyen unas herramientas con capacidades diversas para mostrar a

los usuarios información de la zona de presentación para dar soporte a la toma de decisiones.

Turban presenta un modelo más detallado de un Data Warehouse, este se muestra en la Figura 9. Este modelo presenta adicionalmente los Metadatos que se consideran sumamente importantes en la construcción de un Data Warehouse (Kumar & Reddy, 2006).

Figura 9: Estructura de un Data Warehouse



Fuente: (Turban & Volonino, 2011)

2.3 DATA MART

2.3.1 DEFINICIÓN DE DATA MART

Un Data Mart se define como un repositorio de información, similar a un Data Warehouse, pero orientado a un área o departamento específico de la organización, a diferencia del Data Warehouse que cubre toda la organización, es decir, la diferencia fundamental es su alcance (Kimball, 1998).

Un Data Mart puede también ser definido como una especialización del Data Warehouse, y está enfocado a un departamento o área específica. Permitiendo así un mejor control de la información que se está abarcando (Nader, 2004).

Los Data Mart son un subconjunto de datos del Data Warehouse y es donde tienen lugar la mayoría de las actividades analíticas en el entorno de inteligencia de negocios. Los datos de cada Data Mart son generalmente adaptados para una capacidad particular o función (Imhoff, Galemno, & Geiger, 2003).

Los principales beneficios para una organización de utilizar Data Mart son:

- Acelera las consultas reduciendo el volumen de datos a recorrer.
- Estructura los datos para su adecuado acceso por una herramienta.
- Divide los datos para imponer estrategias de control de acceso.
- Segmenta los datos en diferentes plataformas hardware.
- Permite el acceso a los datos por medio de un gran número de herramientas del mercado, logrando independencia de estas.

Hackney recoge otras definiciones de un Data Mart (Hackney, 1997):

- Meta Group: Un Data Mart es una aplicación orientada a un tema, equivalente a un Data Warehouse, típicamente de un área funcional y está atada a una o más aplicaciones específicas.
- Data Base Associates International: Un Data Mart es una aplicación orientada a un subconjunto de toda la información corporativa, y que los usuarios de un departamento utilizan para tomar mejores decisiones.

- Patricia Seybold Group: Un Data Mart es un recurso de información altamente especializado para resolver un problema específico del negocio.

2.3.2 CARACTERÍSTICAS DE UN DATA MART

En general, los Data Mart tienen las siguientes características (Hackney, 1997):

- Arquitectura empresarial. Se requiere la identificación de áreas organizacionales, dimensiones comunes, fuentes comunes, semántica, métricas y reglas de negocio de la organización.
- Integración. Los Data Mart deben estar integrados con otros Data Mart de la organización.
- Escalabilidad. Los Datmart deben ser creados en plataformas escalables.
- Usuarios homogéneos. Los usuarios deben dar soporte a un grupo con intereses comunes, desafíos u otros requerimientos del negocio similares.

2.3.3 TIPOS DE DATA MART

De acuerdo a la necesidad a la que dan soporte los Data Mart pueden clasificarse en (Imhoff, Galemno, & Geiger, 2003):

1. Data Mart OLAP.
2. Data Mart de exploración.
3. Data Mart estadísticos y de minería de datos.
4. Aplicaciones analíticas personalizables.

Teniendo en cuenta la aproximación del Data Warehouse se pueden dividir en

1. Creados de un DW. Data Mart que son un subconjunto de un DW, reflejaría el enfoque descendente (top-down).
2. Data Mart incremental. Data Mart que tienen fuentes independientes de un DW. Este reflejaría el enfoque ascendente (bottom-up).

2.3.4 TOPOLOGÍA DE UN DATA MART

Las topologías de los Data Mart se dividen según las capas que utilizan (Tabla 4).

Tabla 4: Topologías de un Data Mart

Una Capa	Doble Capa	Tres Capas
BD en una PC	Proceso analítico de bajo nivel.	ROLAP (Proceso analítico en línea relacional).
	Búsqueda y reporte.	MOLAP (Proceso analítico multidimensional).
	Lectores de reportes.	OCSD (Programación y entrega de contenido en línea).
		Intranet.
		Internet.
		Work Flow.
		Colaboración.

Fuente: (Hackney, 1997)

2.3.4.1 TOPOLOGÍA DE UNA CAPA

Este es el caso más pequeño. En este caso el Data Mart completo reside en un pequeño servidor, se trata de una topología típica de "isla de datos". Estos modelos presentan el problema de no mantener la integridad de datos y tienen un riesgo de brindar distintos valores de una misma variable.

2.3.4.2 TOPOLOGÍA DE DOS CAPAS

En este caso el Data Mart está en un servidor distinto al de las fuentes. Este escenario es común para aquellas empresas que tienen la información transaccional en uno o varios servidores centrales.

2.3.4.3 TOPOLOGÍA DE TRES CAPAS

En este caso añade una capa de acceso al Data Mart, mediante un Servidor de Data Mart y/o un Servidor web, y se brinda acceso mediante Intranet/Internet a los datos del Data Mart.

2.3.5 DESARROLLO DE DATA MART

Una metodología para el desarrollo de Data Mart dimensionales se divide en 8 pasos (Chenoweth, Schuff, & Louis, 2003).

1. Analizar el ambiente del usuario desde la perspectiva de toma de decisiones.
2. Desarrollar el modelo lógico, se refieren al esquema de estrella.
3. Seleccionar el sistema administrador de base de datos, ya sea relacional o dimensional.
4. Mapear el sistema lógico en el sistema administrador de base de datos.
5. Desarrollar el modelo físico.
6. Evaluar el modelo físico.
7. Afinar el sistema.
8. Implementar y monitorear el sistema.

2.4 ANÁLISIS DE USO DE UN SITIO WEB

Haciendo una extensión al concepto de minería web de uso de (Escobar, 2007) y (Hu & Cercone, 2004), el análisis de uso un sitio web se puede definir como el descubrimiento de comportamiento de acceso o uso de servicios de un sitio web.

Un término asociado al análisis de un sitio web es el de “analítica web”. Podemos definir a la analítica web como “el cálculo de una variedad de datos, incluyendo el tráfico web, transacciones basadas en web, desempeño de servidores web, estudios de usabilidad, información enviado por usuarios y fuentes relacionadas que ayudan a crear un entendimiento generalizado de la experiencia en línea de un visitante” (Peterson, 2004, p. 6).

Kaushik toma el concepto de la *Web Analytics Association* e indica que “la analítica web es el seguimiento objetivo, recolección, medición, reporte y análisis de los datos cuantitativos de internet para optimizar los sitios web y las iniciativas de marketing en la web” (Kaushik, *Web Analytics: An Hour a Day*, 2007, p. 2).

2.4.1 RECOLECCIÓN DE DATOS PARA ANALIZAR EL USO DE UN SITIO WEB

Existen 4 formas más utilizadas de capturar datos para el análisis de uso de un sitio web (Kaushik, *Web Analytics: An Hour a Day*, 2007):

2.4.1.1 USO DE REGISTROS DEL SERVIDOR WEB

Los registros del servidor han sido la fuente original de análisis de uso de un sitio web. Originalmente se crearon para identificar errores de los servidores web y con el tiempo han ido añadiendo más información, pasando a cubrir de necesidades técnicas a necesidades de marketing.

2.4.1.2 USO DE SERVIDORES RECOLECTORES EXTERNOS (WEB BEACONS)

Los servidores recolectores externos utilizan usualmente imágenes de un pixel por un pixel y transparentes en el servidor web al que se hace seguimiento y con ello crean información para para este último.

2.4.1.3 USO DE ETIQUETAS JAVASCRIPT

En este caso se incluye etiquetas JavaScript en las páginas del servidor web, las mismas envían la información de los accesos a un servidor web –generalmente– externo.

2.4.1.4 USO DE DETECTORES DE PAQUETES

Los detectores de paquetes son herramientas (hardware o software) intermedias entre el usuario y el servidor web. Estos permiten capturar paquetes enviados al servidor web y recolectan la información de los pedidos del visitante.

2.4.2 REGISTROS DE UN SERVIDOR WEB

Los servidores web actualmente permiten el seguimiento de errores y acceso por separado (Apache Software Foundation). El Registro de Errores (*Error Log*)

almacena información sobre registra cualquier error que el servidor web encuentre al procesar peticiones al archivo de registro seleccionado.

El Registro de Accesos almacena la información sobre todas las peticiones que procesa el servidor web. Los formatos más utilizados para los registros de acceso son el Formato Común de Registro (*Common Log Format*) y el Formato de Registro Combinado (*Combined Log Format*).

El Formato Común de Registro tiene la estructura "%h %l %u %t \"%r\" %>s %b", donde:

- (%h) es el host remoto o dirección IP del cliente que hizo la petición al servidor.
- (%l) es la identidad RFC 1413 del cliente determinada por identd en la máquina del cliente. Esta información es poco fiable y no debería ser usada nunca excepto con clientes que estén sometidos a controles muy estrictos en redes internas.
- (%u) es el identificador de usuario de la persona que solicita el documento determinado por la autenticación HTTP. Si el documento no está protegido por contraseña, se mostrará un guión "-" en esta entrada.
- (%t) es la hora a la que el servidor recibió la petición. El formato es "[dd/mmm/aaaa:hh:mm:ss zona_horaria]"
- ("%r") es la línea de la petición del cliente se muestra entre dobles comillas, ésta contiene mucha información de utilidad. Primero, el método usado por el cliente; segundo, el recurso al que ha hecho petición el cliente; y tercero, el protocolo usado por el cliente.
- (%>s) es el código de estado que el servidor envía de vuelta al cliente, ésta revela si la petición fue respondida con éxito por el servidor (códigos que empiezan por 2), una redirección (códigos que empiezan por 3), un error provocado por el cliente (códigos que empiezan por 4), o un error en el servidor (códigos que empiezan por 5).
- (%b) es la última entrada e indica el tamaño del objeto retornado por el cliente sin incluir las cabeceras de respuesta.

El Formato de Registro Combinado añade dos campos al formato previo y la estructura de registros es "%h %l %u %t \"%r\" %>s %b \"%{Referer}i\" \"%{User-agent}i\"", donde:

- ("%{Referer}i") es la cabecera de petición de HTTP "Referer" y muestra el servidor del que proviene el cliente.
- ("%{User-agent}i") es la cabecera de petición HTTP "User-Agent" y brinda información de identificación del agente que usa el cliente.

Los campos adicionales que añade el formato combinado son importantes para evaluar de dónde acceden los clientes de un sitio web.

2.4.3 TIPOS DE DATOS PARA EL ANÁLISIS DE REGISTROS DE UN SITIO WEB

Los datos más utilizados en el análisis de un sitio web capturados a partir de los registros de un servidor son (Ferrini & Mohr, 2009):

- Hit
- Páginas vistas
- Visita (sesión)
- Visitante único
- Visitantes nuevos / visitantes recurrentes
- Páginas vistas por visitante
- Dirección ip
- Localización de un visitante
- Idioma del visitante
- Páginas de referencia / sitios (url)
- Palabras clave
- Navegador
- Sistema operativo
- Resolución de pantalla
- Java o flash activado
- Velocidad de conexión

- Errores
- Duración de la visita
- Rutas de navegación por visitantes
- Porcentaje de rebote

2.4.4 CONSIDERACIONES SOBRE EL ANÁLISIS DE UN SITIO WEB MEDIANTE REGISTROS DE SERVIDORES

Los mayores problemas para el análisis de un sitio web pueden ser generados por (Ferrini & Mohr, 2009) (Kaushik, Web Analytics: An Hour a Day, 2007):

SERVIDORES PROXY

Las páginas servidas por el servidor proxy no se registran en los archivos de registro del servidor web, dando lugar a imprecisiones en el conteo de tráfico del sitio. Esta situación puede hacer que se presenten subestimaciones en el tráfico de visitantes únicos.

Por otro lado, a veces a los usuarios domésticos con un proveedor de servicios de Internet se les asigna una nueva dirección IP cada vez que se conectan a Internet. Esto provoca el efecto contrario de inflar el número de visitas únicas en los registros Web.

CORTAFUEGOS (FIREWALLS)

Los *cortafuegos (firewall)* ocultan la dirección IP real para equipos específicos del usuario y en su lugar presentan una única dirección IP genérica a Internet para todos los usuarios. Por lo tanto, esto hace que se presenten subestimaciones en el tráfico de visitantes únicos en el sitio web.

ALMACENAMIENTO EN CACHE

En lugar de solicitar la misma página de nuevo desde el servidor (por ejemplo, si el usuario hace clic en el botón "atrás"), el navegador de un ordenador mostrará una copia de la página en lugar de hacer otra nueva petición al servidor.

Al igual que con el uso de servidores proxy, el almacenamiento en caché plantea un problema dado que las páginas vistas almacenadas en caché no son registradas en los registros del servidor web.

ORIENTACIÓN A CAPTURA DE ERRORES

Los registros web se orientaron en un principio a la captura de información técnica (404 errores, tendencias de uso del servidor, tipo de navegador, etc) que no son adecuados para capturar la información comercial.

DEPENDENCIA DEL ÁREA DE TECNOLOGÍAS DE INFORMACIÓN (TI)

Si se requiere información adicional, la nueva captura de datos requiere una estrecha colaboración del equipo de administración de contenidos y el área de TI.

CÁLCULOS COMPLEJOS

El análisis de registros web requiere utilización adecuada de filtros para eliminar solicitudes de imágenes, errores de páginas, tráfico robot, solicitudes de otros elementos, con el fin de conseguir tráfico preciso.

2.4.5 BENEFICIOS DE UTILIZAR REGISTROS DE SERVIDORES PARA EL ANÁLISIS DE USO DE UN SITIO WEB

Los mayores beneficios de utilizar los registros web son (Kaushik, Web Analytics: An Hour a Day, 2007) (Jansen, Taksa, & Spink, 2009) (Peterson, 2004):

ACCESIBILIDAD DE DATOS

Cada servidor web viene con mecanismos que recogen los datos y crean registros Web de las solicitudes al sitio web.

HERRAMIENTAS GRATUITAS

Hay muchos programas de análisis de archivos de registros web disponibles de forma gratuita, esto permite crear informes rápidamente.

IDENTIFICACIÓN DE COMPORTAMIENTO DE ROBOTS

Los registros del servidor permiten analizar el comportamiento de los robots de los motores de búsqueda (y otros tipos) en un sitio web. Robots de motores de búsqueda no dejan rastro en otros mecanismos de captura de datos.

PROPIEDAD DE LOS DATOS

Este mecanismo evita la salida de datos de la organización y permite además independencia de la herramienta de análisis que se utilice o del modelo utilizado. Con la mayoría de otras metodologías, los datos serán capturados, procesados y almacenados por el proveedor de analítica web.

El manejo de la privacidad depende únicamente de la organización.

FLEXIBILIDAD DE LA RECOLECCIÓN DE DATOS

Los servidores web pueden programados para almacenar ciertos tipos de datos e ignorar otros.

IDENTIFICACIÓN DE ESTADO DE DESCARGAS

La información de bytes descargados, así como la de identificación del estado de los archivos solicitados por los visitantes puede ser capturado por los registros del servidor.

MÉTODO DISCRETO, NO INTRUSIVO

Una de las principales características es que se trata de un método no intrusivo, es decir, no requiere una interacción directa con los participantes (visitantes) en la investigación (u obtención de datos).

IDENTIFICACIÓN DE TODAS LAS TRANSACCIONES CON EL SERVIDOR

El uso de registros de servidor garantiza la utilización de todos los registros provenientes de todas las transacciones en el sitio web.

Los métodos alternativos que utilizan etiquetas JavaScript no identifican el acceso de usuarios que usan navegadores con JavaScript deshabilitado. Estos métodos tampoco pueden capturar descargas de documentos (PDF, XLS, DOC, etc) a partir de enlaces externos al sitio web en análisis.

2.5 ENTIDADES PÚBLICAS

2.5.1 DEFINICIÓN

De acuerdo a la Directiva N°002-2010-PCM/SGP (Presidencia del Consejo de Ministros, 2010):

“Se considera Entidad Pública a toda organización del Estado Peruano, con Personería jurídica de Derecho Público, creada por norma expresa en el que se le confiere mandato a través del cual ejerce funciones dentro del marco de sus competencias y atribuciones, mediante la administración de recursos públicos, para contribuir a la satisfacción de las necesidades y expectativas de la sociedad, y como tal está sujeta al control, fiscalización y rendición de cuentas”.

2.5.2 PORTAL DE INTERNET DE UNA ENTIDAD PÚBLICA

Para esta investigación se considerará como Portal de Internet de una entidad pública al sitio web principal que tiene la información institucional y de transparencia. Además que tiene un responsable definido por Ley y está indexado en directorio del Portal del Estado Peruano.

2.5.3 PORTAL DEL ESTADO PERUANO (PEP)

El PEP tiene la siguiente dirección <http://www.peru.gob.pe>

Mediante el Decreto Supremo N° 060-2001-PCM se crea el "Portal del Estado Peruano" como Sistema Interactivo de Información a los ciudadanos a través de Internet (Presidencia del Consejo de Ministros, 2001).

El PEP se crea como un Portal de información que proporciona un servicio de acceso unificado sobre los servicios y procedimientos administrativos que se realizan ante las diversas dependencias públicas.

El PEP tiene la máxima jerarquía entre los Portales, páginas web y similares de la Administración Pública (Presidencia del Consejo de Ministros, 2004), esto implica que los contenidos de todos los sitios web de las entidades públicas deben estar conectados o enlazados al PEP.

La información contenida en el Portal del Estado Peruano y en los Portales Institucionales tiene carácter y valor oficial. Además la responsabilidad de publicación de los documentos normados por las leyes peruanas recae sobre los funcionarios responsables de la elaboración de los portales de internet de cada entidad pública (Congreso de la República, 2007).

El Reglamento de la Ley N° 29091 especifica los documentos que se deben publicar en los Portales Instituciones de las entidades del Estado, en el Portal del Estado Peruano y en el Portal de Servicios al Ciudadano y Empresas (Presidencia del Consejo de Ministros, 2008).

2.5.4 PORTAL DE SERVICIOS AL CIUDADANO Y EMPRESAS (PSCE)

El PSCE tiene la siguiente dirección <http://www.serviciosalciudadano.gob.pe/>

El artículo 2 del Decreto Supremo N° 032-2006-PCM crea al Portal de Servicios al Ciudadano y Empresas "adscrito al Portal del Estado Peruano – PEP, como un sistema de información en Internet sobre los procedimientos, requisitos y derechos establecidos en el Texto Único de Procedimientos Administrativos - TUPA de las entidades públicas del Estado, a fin de facilitarle dicha información al administrado, ..." (Presidencia del Consejo de Ministros, 2006).

2.5.5 TRANSPARENCIA Y ACCESO A LA INFORMACIÓN

La Ley Transparencia y Acceso a la Información Pública indica el “Principio de publicidad” (Congreso de la República, 2002) ello implica:

- Toda información que posea el Estado se presume pública, salvo las excepciones expresamente previstas en esta Ley.
- El Estado adopta medidas básicas que garanticen y promuevan la transparencia en la actuación de las entidades de la Administración Pública.
- El Estado tiene la obligación de entregar la información que demanden las personas en aplicación del principio de publicidad.

Además esta Ley indica que todas las entidades públicas designarán al funcionario responsable de entregar la información solicitada.

Esta Ley establece la creación de portales de transparencia en las entidades del Estado y también indica que las entidades públicas deberán identificar al funcionario responsable de la elaboración de los portales de Internet.

2.5.6 PORTAL DE TRANSPARENCIA ESTÁNDAR

El Portal de Transparencia estándar principal se encuentra dentro del Portal del Estado peruano y además se incorpora un enlace en las páginas web de las entidades públicas.

El Decreto Supremo N° 063-2010-PCM aprueba la implementación del Portal de Transparencia Estándar en las entidades públicas (Presidencia del Consejo de Ministros, 2010).

Este Decreto establece formatos estándares bajo los cuales las entidades públicas están obligadas a publicar la información de la Ley Transparencia y Acceso a la Información Pública.

2.5.7 ACCESIBILIDAD DE LOS SITIOS WEB DE ENTIDADES PÚBLICAS

El gobierno promueve la accesibilidad de los sitios web y obliga a que las entidades públicas adecuen sus portales y páginas para que personas con discapacidad visual puedan acceder a la información que contienen estos sitios web (Congreso de la República, 2005).

La Resolución Ministerial N° 126-2009-PCM establece los lineamientos para accesibilidad a páginas web y aplicaciones para telefonía móvil para instituciones públicas del Sistema Nacional de Informática (Presidencia del Consejo de Ministros, 2009).

Con esta Resolución el Estado Peruano establece el propósito de que las instituciones públicas apliquen las "Pautas de Accesibilidad de los Contenidos Web 1.0". El documento se basa en las directivas de la W3C (*World Wide Web Consortium*, es un consorcio internacional que produce estándares para la *World Wide Web*).

CAPÍTULO 3

ESTADO DEL ARTE

3.1 CLASIFICACIÓN DE LA INVESTIGACIÓN

De acuerdo a la clasificación propuesta por (Barrantes, 1999) esta investigación se clasificaría según la Tabla 5.

Tabla 5: Clasificación de la Tesis

Clasificación según:	Tipo
Finalidad	Investigación aplicada: Busca solucionar un problema práctico para los administradores de contenido de sitios web de entidades públicas.
Alcance temporal	Estudios transversales: Estudia la aprobación del modelo durante el periodo de la investigación.
Profundidad u objetivo	Descriptiva: Usa la observación, estudios correlacionados para describir un escenario base del modelo SSD-IWEP.
Carácter de la medida	Cuantitativo: Se mide la aprobación de los administradores de contenido de sitios web de entidades públicas.
Marco en que tiene lugar	De campo: La investigación se realiza en situaciones naturales y esto permite generalizar los resultados a situaciones similares.

Concepción del fenómeno	Nomotética: Busca establecer conclusiones, la investigación usa la metodología empírica y analítica y se apoya en la experimentación.
Dimensión temporal	Experimental: La investigación incorpora una encuesta orientada a desarrollar aplicaciones basadas en el modelo en el futuro.
Orientación que asume	Orientada a la explicación: Busca dar respuesta a problemas concretos para la toma de decisiones.

Fuente: (elaboración propia).

3.2 ANTECEDENTES

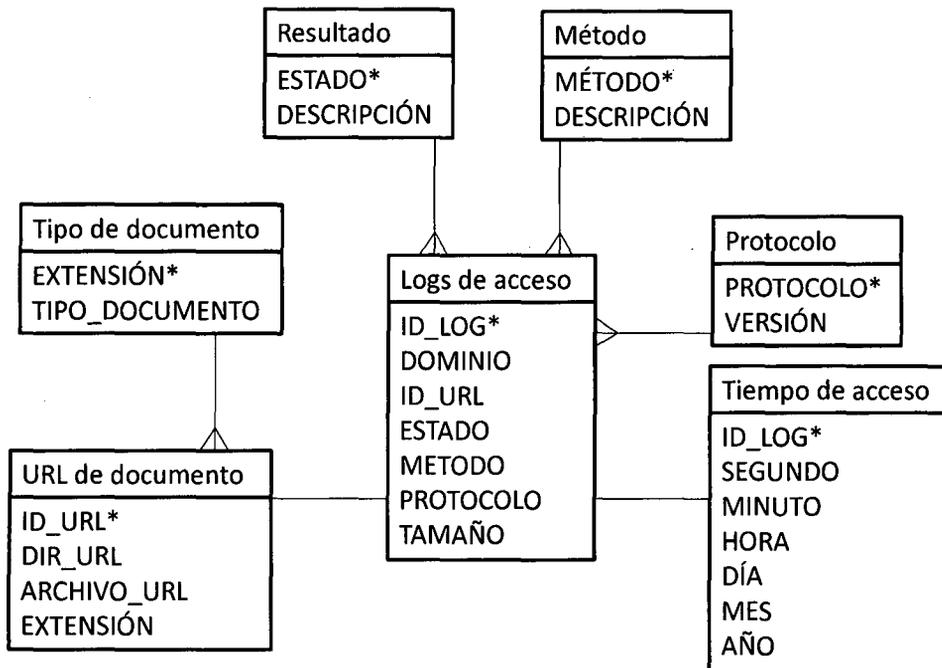
Los sistemas informáticos de apoyo a la toma de decisiones se volvieron viables con el desarrollo de las minicomputadoras, sistemas operativos compartidos y sistemas distribuidos (Power D. , Decision Support Systems: A Historical Overview, 2008b). En los 60s, empiezan los estudios sobre modelos cuantitativos computarizados para apoyar en la toma de decisiones.

3.2.1 NORMALIZACIÓN DE LOS REGISTROS DE UNA PÁGINA WEB

La Figura 10 muestra una base de datos normalizada de los registros de una página web (Joshi, Joshi, Yesha, & Krishnapuram, 1999). Este almacenamiento en bases de datos incrementa la seguridad de los registros del servidor web, reduce el espacio de almacenamiento, aumenta la accesibilidad y permite realizar consultas manera más eficiente.

El análisis de los registros del servidor web no sólo brinda información importante a los administradores de contenido, también ayuda para poder crear páginas web de contenido adaptativo.

Figura 10: Base de datos normalizada para los registros de la página web



Fuente: (Joshi, Joshi, Yesha, & Krishnapuram, 1999)

3.2.2 UN DATA MART PARA EL ANÁLISIS DE HITS

Estos autores (Joshi, Joshi, Yesha, & Krishnapuram, 1999) presentan además un esquema para un Data Mart (Figura 11), cuya tabla de hechos se formularía a través de una conversión de la tabla normalizada, ellos plantean tener las siguientes dimensiones:

- Nombre del dominio.
- Fecha y hora.
- Método de la petición.
- URL de la página accedida.
- Resultado del pedido.
- Número de bytes transmitidos.

Figura 11: Tabla de hechos

Tabla de hechos
ID_LOG*
DOMINIO
DIA
MES
AÑO
HORA
MINUTO
SEGUNDO
METODO
URL
PROTOCOLO
ESTADO
TAMAÑO

Fuente: (Joshi, Joshi, Yesha, & Krishnapuram, 1999)

El modelo previo no muestra explícitamente la actividad de las visitas. Para poder mostrar las visitas se requiere diseñar un algoritmo que permita calcular a partir de hits las vistas, las visitas y los visitantes. Estos cálculos son bastante complejos de realizar por distintas razones como la presencia de proveedores de internet (ISP), distintos niveles de caché, datos incompletos y abundancia de información en los ficheros de registros.

Otro problema en esta situación lo añaden los robots, spiders y agentes inteligentes que reúnen información para sus bases de datos y generan ruido en los registros del servidor web.

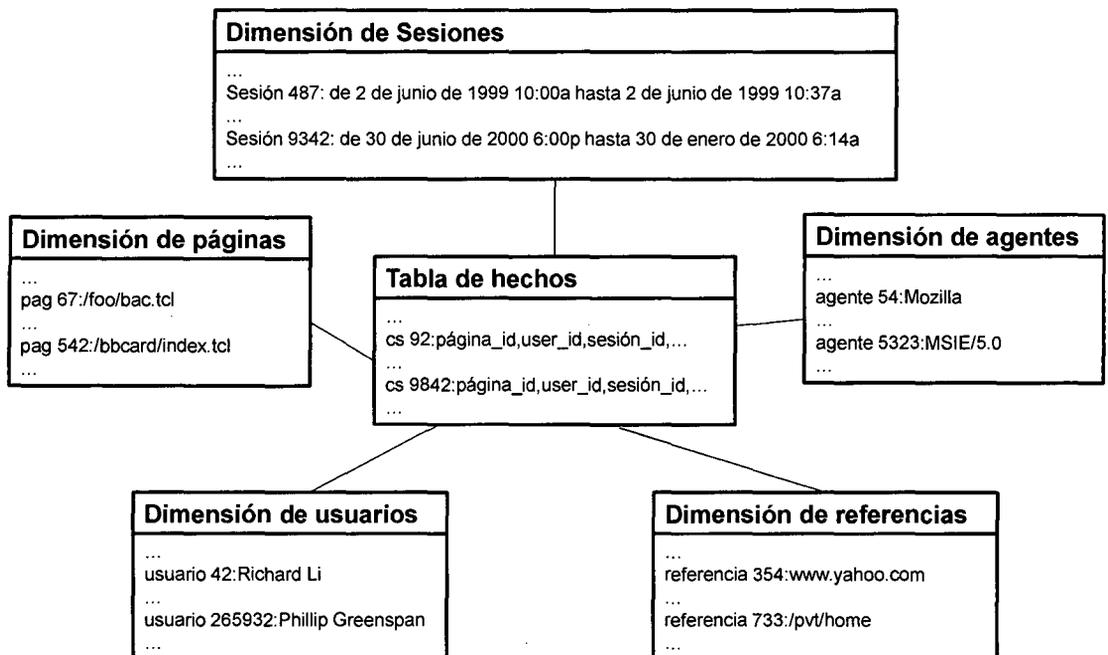
Una vez superado la identificación de visitantes, visitas y páginas, el siguiente paso es crear un modelo de análisis.

3.2.3 ANÁLISIS DE USO DE UNA PÁGINAS WEB

La propuesta de Data Mart con 5 dimensiones (Li, 2000) que se muestra en la Figura 12 se basa en el seguimiento de clics. Este modelo plantea estas dimensiones.

- Sesiones.
- Páginas.
- Agentes.
- Usuarios.
- Referencias.

Figura 12: Dimensiones del modelo de datos de seguimiento de clics



Fuente: (Li, 2000).

Se debe destacar que se deben tener en cuenta 5 características fundamentales en el diseño de su modelo (Li, 2000):

- Privacidad de los visitantes.
- Desempeño y escalabilidad.
- Transparencia.
- Independencia de plataforma.
- Extensibilidad.
- Exactitud.

Este también puede ser utilizado para un modelo basado en análisis de los registros de un servidor web.

3.2.4 ANÁLISIS DE SITIOS WEB SEGÚN USUARIO

El análisis de estadísticas de uso, según a quien vaya orientado, se puede dividir en dos partes:

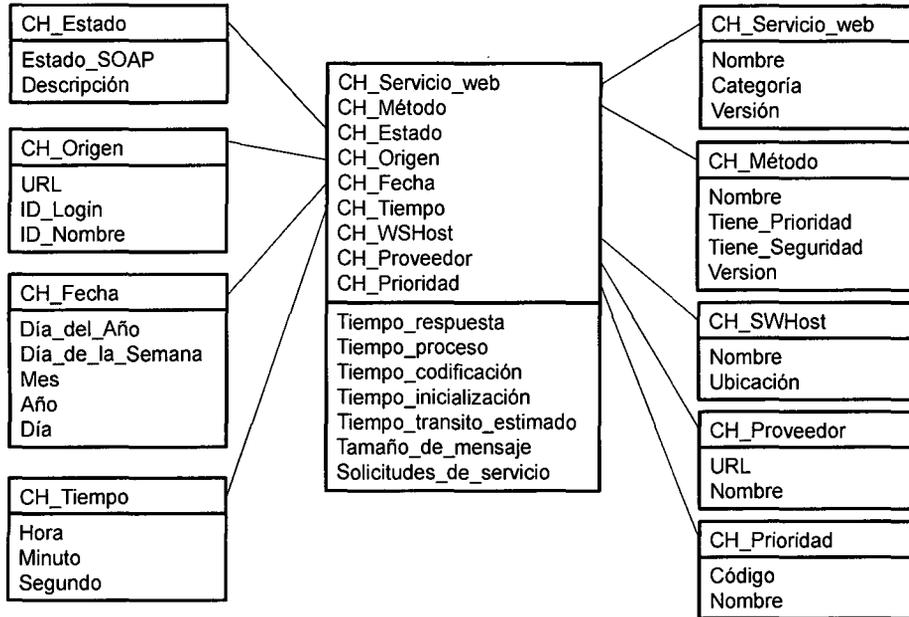
- Orientado a los administradores de la página web
- Orientado al análisis de ventas

Un modelo con 2 Data Mart (Serra, Campos, Machado, & Pires, 2003)³ permite separar la información que necesitan ambos lados de la organización. El primer Data Mart planteado por estos autores incluye 9 dimensiones (Figura 13) y está orientado al servicio web:

- Estado
- Origen
- Fecha
- Hora
- Servicio web
- Método
- Host
- Proveedor
- Prioridad

³ Serra, S., Campos, L., Machado, M. y Pires, P. "A data mart approach for monitoring Web services usage and evaluating quality of services" en el XVIII Simpósio Brasileiro de Banco de Dados. 2003.

Figura 13: Data Mart para monitorear el servidor web

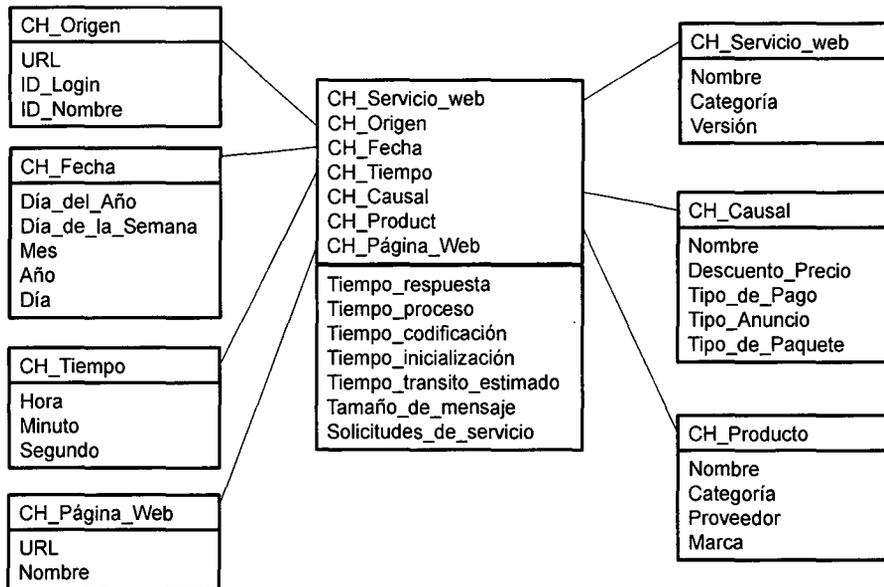


Fuente: (Serra, Campos, Machado, & Pires, 2003)

El segundo Data Mart orientado al análisis del producto (Figura 14) incluye 7 dimensiones:

- Origen.
- Fecha.
- Hora.
- Página web.
- Servicio web.
- Características de la venta.
- Producto.

Figura 14: Data Mart para los productos



Fuente: (Serra, Campos, Machado, & Pires, 2003)

Este segundo Data Mart planteado (Serra, Campos, Machado, & Pires, 2003) considera una organización orientada a la venta de productos, sin embargo se podría redirigir a un enfoque de difusión de información si consideramos que cada publicación es un producto en sí.

3.2.5 MODELO PARA ANÁLISIS DE VISITAS Y PÁGINAS CON DOS ESTRELLAS

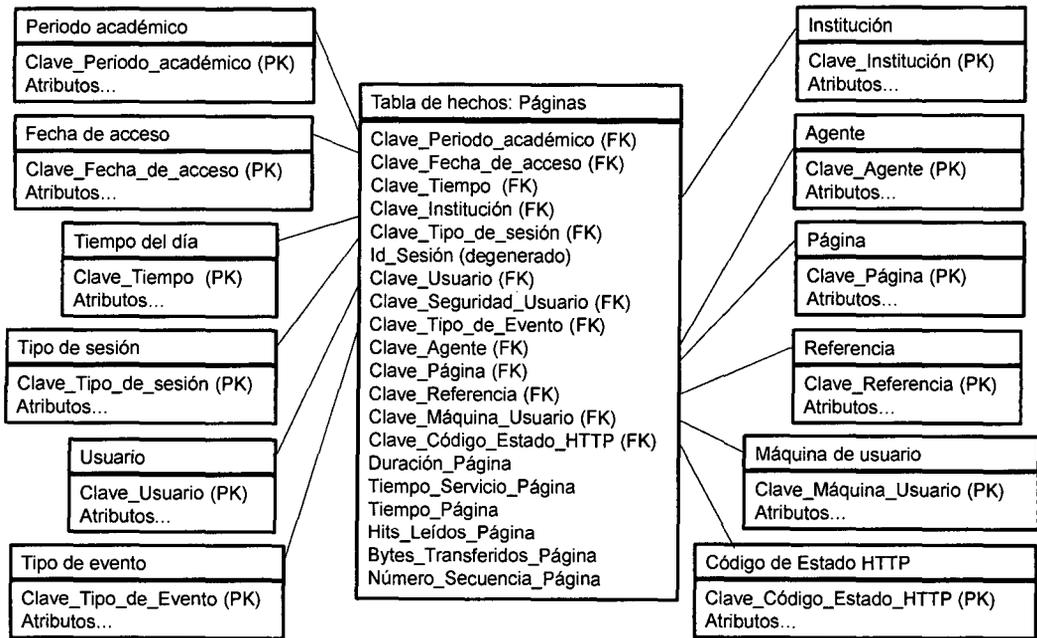
Una forma de dividir el análisis de uso de un sitio web es separarlo por análisis de usuarios y análisis de páginas.

Un modelo utilizado en la *Universidade do Porto* (Teixeira & David, 2005) presentan un modelo para monitorear la conducta de los usuarios de un sistema de información para evaluar su impacto y orientar su mejora, para ello plantean dos tablas de hechos, una orientada al análisis de uso de las páginas (Figura 15) y la otra dirigida a evaluar el comportamiento de las sesiones (Figura 16).

La estructura de los Data Mart nos muestra una estructura propia de la Universidad, como el periodo académico y la institución, esto indica una adecuación de acuerdo a las necesidades de información de la institución.

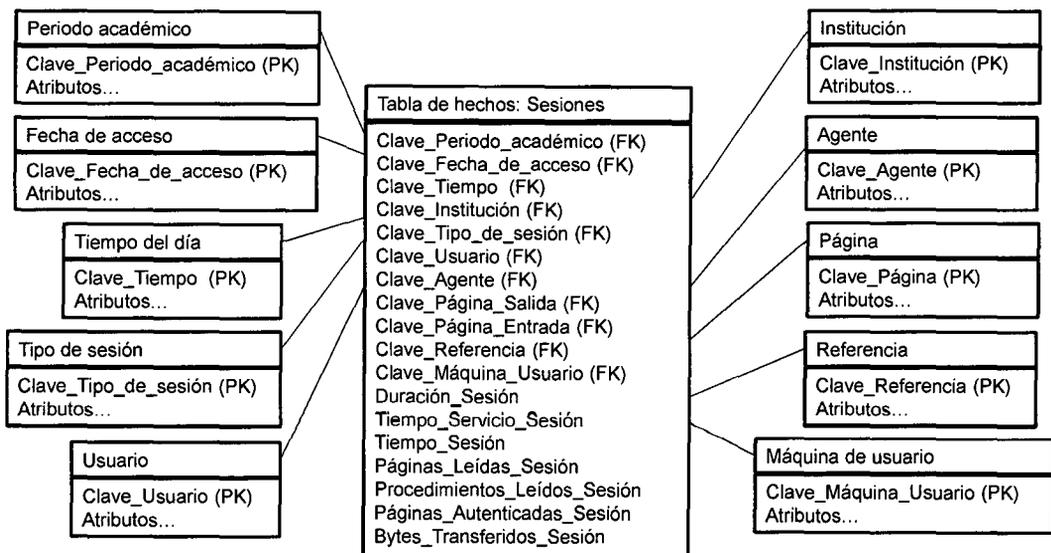
Este modelo es importante porque añade la posibilidad de añadir agrupaciones o características particulares que escapan a las características típicas de los registros del servidor.

Figura 15: Data Mart para análisis de páginas



Fuente: (Teixeira & David, 2005)

Figura 16: Data Mart para análisis de sesiones

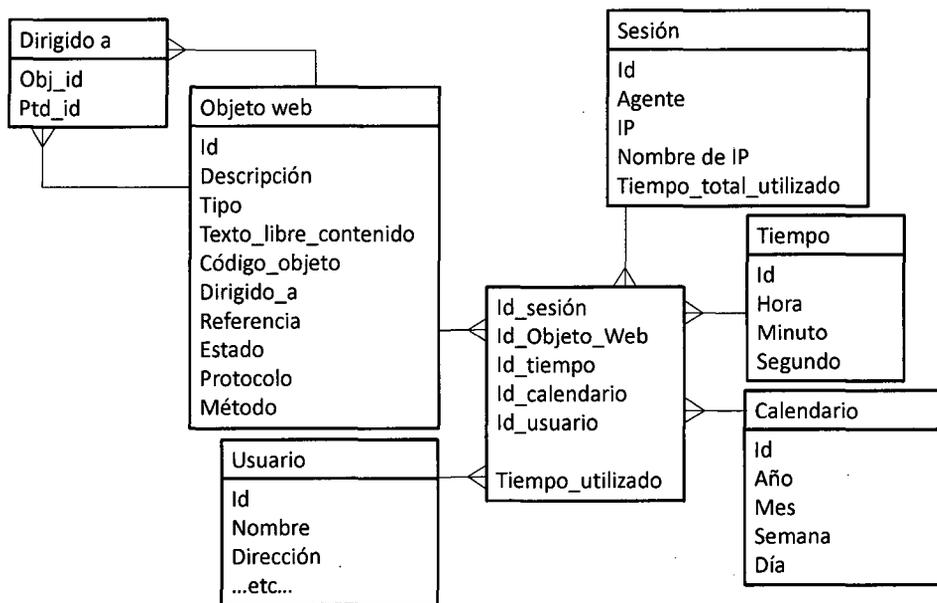


Fuente: (Teixeira & David, 2005)

3.2.6 MODELO GENÉRICO PARA ANÁLISIS DE VISITAS Y PÁGINAS

El modelo genérico planteado por (Velásquez & Palade, 2008) presenta un modelo alternativo de un Data Mart con 5 dimensiones. Estos autores indican que, en términos de desempeño, no es conveniente definir un modelo de estrella con muchas tablas de dimensión debido a que las operaciones que incluyen unir tablas (JOIN) utilizarían sustanciales recursos computacionales.

Figura 17: Un modelo genérico de estrella para los datos de la página web

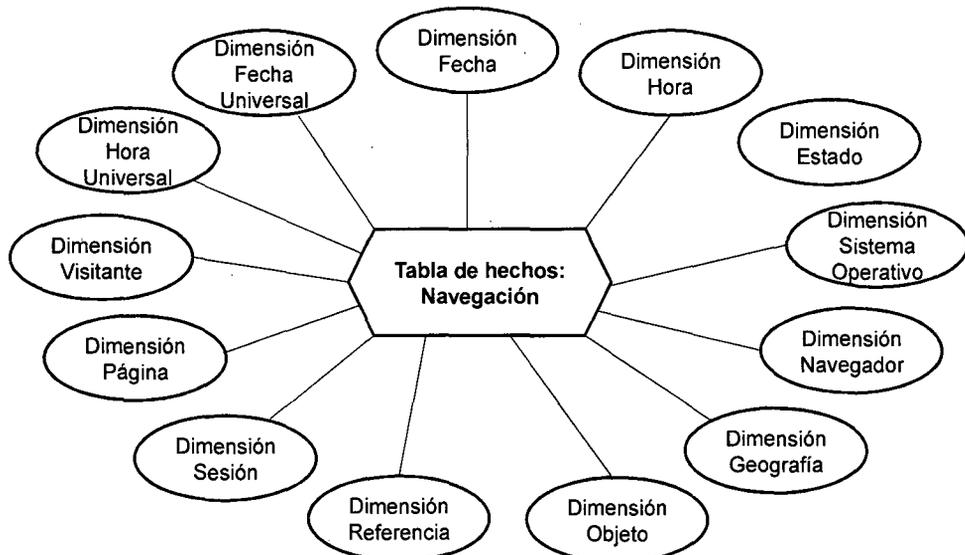


Fuente: (Velásquez & Palade, 2008).

3.2.7 ESQUEMA DE ESTRELLA DETALLADO PARA LA NAVEGACIÓN

Por otro lado, (Ravikumar, Manjunath, Ravindra, & Archana, 2011) presentan un esquema de estrella de Data Mart más general para el seguimiento de clics (Figura 18).

Figura 18: Un modelo de estrella de Data Mart orientado a páginas web



Fuente: (Ravikumar, Manjunath, Ravindra, & Archana, 2011)

En el centro la tabla de hechos básicamente que recoge el tiempo en segundos que demora un visitante en una página o archivo y los bytes transferidos al cliente del servidor web. Las dimensiones están descritas a continuación:

- Fecha universal con niveles de Año, Trimestre, Mes, Semana y Día.
- Fecha universal del día con niveles de Hora, Minuto y Segundo.
- Hora local con niveles de Año, Trimestre, Mes, Semana y Día.
- Hora Local del día con niveles de Hora, Minuto y Segundo.
- Dirección IP, id de visitante o id de cookie.
- Página con niveles de tipo de objeto, tipo de archive, página y dirección URL o Dominio, sitio, directorio y URL.
- Sesión con nivel de sesión (visita) y tipo.
- Referencia con niveles de tipo y URL o Dominio, Sitio y URL.
- Estado con niveles de tipo y descripción.
- Visita.
- Página de contenido.

3.3 BASE DEL MODELO SSD BUSCADO

La investigación ampliará el modelo de normalización de los registros planteado previamente (Joshi, Joshi, Yesha, & Krishnapuram, 1999) y a partir de este planteará un modelo de análisis de datos para el SSD-IWEP basado en Data Mart con 2 estrellas (Teixeira & David, 2005), una para el análisis de sesiones y otra para el análisis de páginas. El modelo SSD-IWEP extenderá una estrella para análisis de visitantes.

La investigación evalúa todas las dimensiones planteadas (Ravikumar, Manjunath, Ravindra, & Archana, 2011) y plantea un modelo

3.4 APLICACIONES SIMILARES

En esta sección se toman en cuenta las herramientas gratuitas más utilizadas por los administradores de sitios web. AWStats, Analog y Google Analytics son tres de las herramientas de análisis de registros web gratuitos más utilizadas.

3.4.1 AWSTATS

AWStats, separa a los visitantes humanos de los robots de búsqueda en su informe resumido de tráfico. El comportamiento de los robots web, arañas y rastreadores es algo que las herramientas basadas en JavaScript no pueden mostrar, porque los motores de búsqueda no ejecutan JavaScript y por ende no pueden enviar los datos al servidor de rastreo.

Además las herramientas basadas en análisis de registros pueden mostrar los bytes consumidos y los errores de páginas no encontradas (King, 2008).

AWStats genera unos análisis bastante robustos, puede analizar registros de servidores: Web, de contenido multimedia a demanda, de transferencia de archivos y de correo. También se puede analizar archivos de registro de todas las herramientas de servidor más importantes. Esta herramienta de código abierto es muy fácil de leer y ofrece una rápida lista de estadísticas globales. Incluso, pueden obtener información sobre los ataques de gusanos con este software. Esta herramienta gratuita funciona con cualquier sistema operativo o navegador.

3.4.2 ANALOG

Analog es otro analizador de archivo de registro muy popular. Ofrece varias ventajas. Es escalable, altamente configurable y gratis.

Funciona en cualquier sistema operativo y navegador, y se reporta en 32 idiomas. Esta herramienta gratuita ofrece apoyo o soporte comercial. Es muy fácil de leer y ofrece una rápida ronda de estadísticas globales.

3.4.3 GOOGLE ANALYTICS

Google ofrece una herramienta de análisis gratuita y brinda abundante información acerca de cómo los visitantes encuentran un sitio web y cómo interactúan con él.

Google Analytics presenta limitaciones debido al uso de etiquetas JavaScript que no permiten analizar algunos aspectos importantes en sitios web educativos o por las definiciones de visita preestablecidos (Dragos, 2011).

3.5 INDICADORES UTILIZADOS PARA EL ANÁLISIS DE UN SITIO WEB

Un producto de analítica web debe permitir obtener los siguientes indicadores (Kaushik, 2010) (Peterson, 2004):

- Visitas
- Visitantes únicos
- Descargas
- Tiempo en la página
- Tiempo en el sitio web
- Ratio de rebote
- Páginas de salida
- Páginas de entrada
- Referencias
- Tasas de conversión
- Frecuencia de uso

CAPÍTULO 4

MODELO SSD-IWEP

4.1 ANÁLISIS INICIAL

4.1.1 OBJETIVO

Desarrollar la primera versión de un Modelo SSD orientado a la administración de contenido de sitios web de entidades públicas en el Perú (SSD-IWEP), que permita dar soporte a las decisiones de los administradores de contenido de estos portales de Internet.

El modelo debe permitir la construcción de aplicaciones que brinden información sobre:

- ¿Cuál es el comportamiento de los visitantes al sitio web?
- ¿Identificar el comportamiento de las visitas en el sitio web?
- ¿Cuáles son las características de las descargas al sitio web?

Esta información permitirá dar soporte para tomar mejores decisiones sobre:

- Información a difundir.
- Información orientada a distintos públicos.
- Contenido a publicar o retirar del sitio.
- Promover el sitio en otros sitios web.
- Estructura del sitio web.
- Identificación y manejo de robots.

- Publicación de documentos.

El modelo debe tener las siguientes características:

- No intrusivo
- Seguro
- Mantiene la privacidad
- Desempeño
- Escalabilidad
- Transparente en diseño
- Independiente de la herramienta
- Extensible
- Exacto
- Accesible
- Filtra robots
- Mantiene la propiedad
- Identifica todas las transacciones

4.1.2 LÍMITES DEL MODELO

Los límites del modelo son los siguientes:

- El modelo se centra en el análisis de los archivos de registros (log) del servidor web.
- El modelo no propone modelos matemáticos, ni estadísticos para el análisis de visitantes y descargas.
- El modelo asume que el principal producto de estos sitios web es la información que brindan.
- El modelo está preparado para explotación de datos, sin embargo no incluye el manejo de modelos de minería de datos (*Data Mining*).

4.2 ANÁLISIS SITUACIONAL

4.2.1 MODELO DE FUENTE DE DATOS

La fuente de datos del sistema lo conforman los siguientes componentes principales:

4.2.1.1 DATOS DE LA ESTRUCTURA DEL SITIO WEB

La estructura de la información del sitio web dependerá de cada entidad pública, esta debe tener al menos la siguiente tabla:

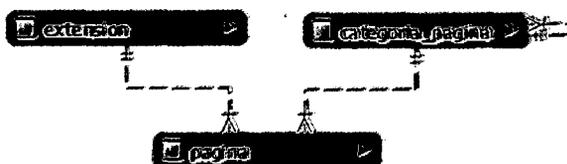
- Tabla “paginas”: La información básica de las páginas web, esta incluye el identificador de página y el nombre.

Adicionalmente el modelo plantea la categorización de las páginas y palabras claves, el modelo SSD-IWEP para ello incorpora las siguientes tablas:

- Tabla “categoria”: Esta tabla guarda las categorías de las páginas. Esta tabla requiere que la tabla “paginas”.
- Tabla “extension”: Esta tabla almacena las extensiones de las páginas.

La Figura 19 muestra la relación de las tablas asociadas a la estructura del sitio web.

Figura 19: Tablas para la estructura del sitio



Fuente: (Elaboración propia).

4.2.1.2 DATOS PROVENIENTES DE LA IP

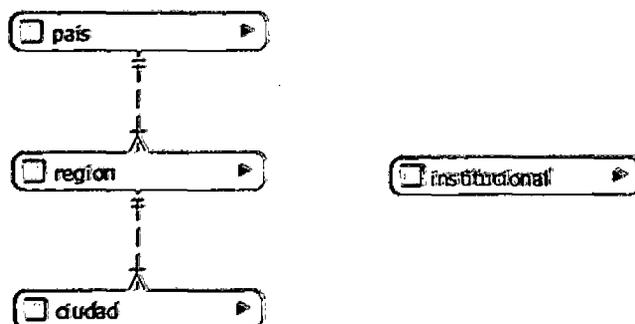
El modelo plantea la utilización de un servicio de identificación geográfica (GeoIP) e identificación de empresas. Los servicios actuales de GeoIP permiten conocer con cierta certeza el origen de las consultas basándose en las IP de los visitantes,

además se puede utilizar este para identificar a los visitantes institucionales. Este grupo estaría conformado por las siguientes tablas:

- Tabla "ciudad": Esta guarda la información de las ciudades del planeta y estaría asociado a una región.
- Tabla "region": Esta tabla guarda datos de la región y estaría asociado de un país
- Tabla "pais": Almacena datos de los países actuales en el planeta.
- Tabla "institucional": Guardaría la información de los visitantes institucionales.

La Figura 20 muestra la relación de las tablas asociadas a la IP de los visitantes.

Figura 20: Tablas asociadas a la IP



Fuente: (Elaboración propia).

4.2.1.3 DATOS PROVENIENTES DEL AGENTE DE NAVEGACIÓN

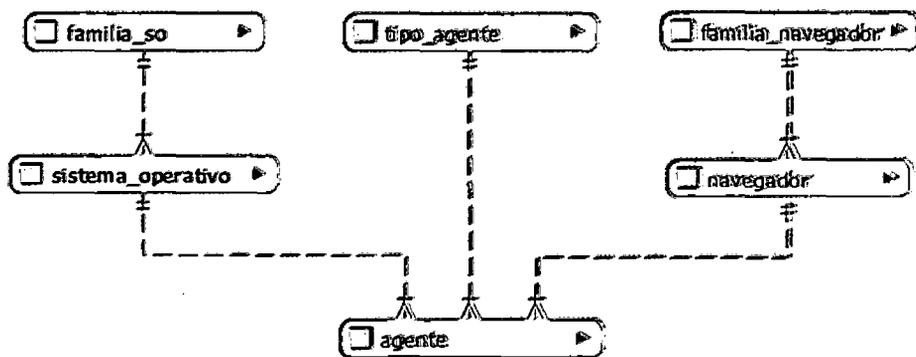
El modelo plantea usar un servicio externo de identificación de agentes de navegación. Este servicio permite conocer el navegador y el sistema operativo de los visitantes a un sitio web. Las tablas requeridas para este servicio son:

- Tabla "agente": Almacena la información del identificador del agente y su descripción.
- Tabla "tipo_agente": Brinda información sobre el tipo de agente utilizado en la sesión.

- Tabla “sistema_operativo”: Guarda la información del sistema operativo utilizado por el visitante.
- Tabla “familia_so”: Guarda la información de las familias de sistema operativo utilizado por el visitante.
- Tabla “navegador”: Esta tabla almacena los navegadores utilizados por los visitantes.
- Tabla “familia_navegador”: Esta tabla guarda las familias de navegadores utilizados por los visitantes.

La Figura 21 muestra la relación de las tablas asociadas al agente de navegación.

Figura 21: Tablas asociadas al agente de navegación



Fuente: (Elaboración propia).

4.2.1.4 DATOS PROVENIENTES DE LOS REGISTROS DEL SERVIDOR WEB

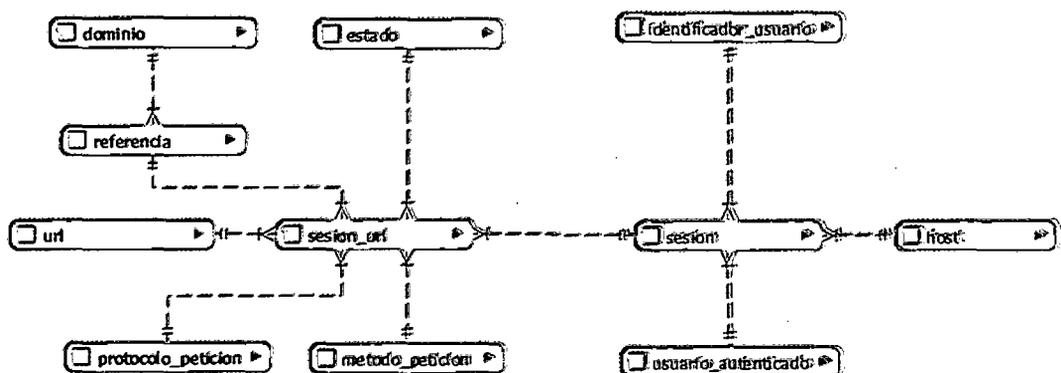
El modelo SSD-IWEP extiende la investigación planteada en la sección 3.2.1 (Joshi, Joshi, Yesha, & Krishnapuram, 1999). El Modelo SSD-IWEP plantea las siguientes tablas:

- Tabla “sesion”: Esta tabla almacena información de las sesiones iniciadas en el servidor.
- Tabla “host”: Almacena la dirección IP del visitante.
- Tabla “identificador_usuario”: Almacena la identidad RFC 1413 del cliente.
- Tabla “usuario_autenticado”: Guarda los identificadores de usuarios autenticados.
- Tabla “url”: Almacena la información de las direcciones URL utilizadas.

- Tabla "sesión_url": Almacena las URL visitadas.
- Tabla "estado": Guarda información sobre el estado de la solicitud de la URL.
- Tabla "protocolo_petición": Guarda el protocolo utilizado en la petición de la información.
- Tabla "metodo_petición": Almacena el método de petición.
- Tabla "referencia": Guarda las referencias de las URL.
- Tabla "dominio": Guarda los dominios que hacen referencia al sitio web.

La Figura 22 muestra la relación de las tablas necesarias para normalizar los datos provenientes de los registros del servidor Web.

Figura 22: Tablas provenientes de los registros del servidor web



Fuente: (Elaboración propia).

4.2.2 ANÁLISIS DE REQUERIMIENTOS

Los reportes que ayudan a tomar decisiones sobre la difusión de información en un sitio web son:

- Reporte regional con desagregación de País, Región y Ciudad.
- Reporte de visitantes institucionales.
- Reporte de páginas con agrupación de categorías.
- Reporte de páginas de entrada con agrupación de categorías.
- Reporte de páginas de salida con agrupación de categorías.
- Reporte de páginas de referencia con agrupación de dominios.
- Reporte de sistemas operativos con desagregación de versiones.

- Reporte de navegadores con desagregación de versiones.
- Reporte de actividad por horas.

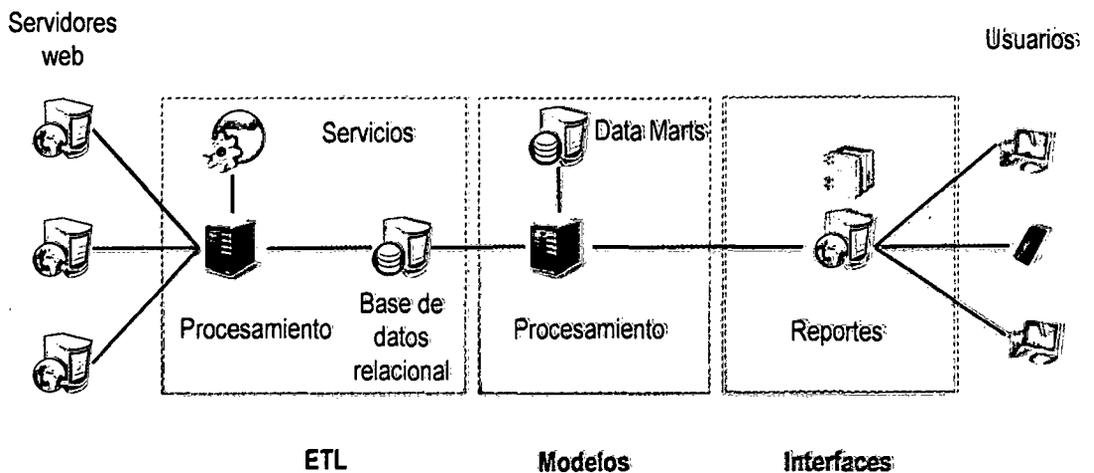
Todos los reportes deben tener:

- Filtro de navegación
 - Visitantes únicos
 - Visitas
 - Descargas
- Filtro por tipo de visitante
 - Total
 - No Robots (visitantes humanos)
 - Robots
- Manejo de periodos de consulta
 - Definición de fecha inicial
 - Definición de fecha final
 - Definición de agrupación por meses, semanas o días

4.2.3 DIAGRAMA GENERAL DEL SSD-IWEP

La Figura 23 muestra el diagrama general del Modelo SSD-IWEP, en esta se puede mostrar los 3 componentes: Datos, Modelos de análisis e Interfaces de acceso.

Figura 23: Diagrama general del modelo SSD-IWEP



Fuente: (Elaboración propia).

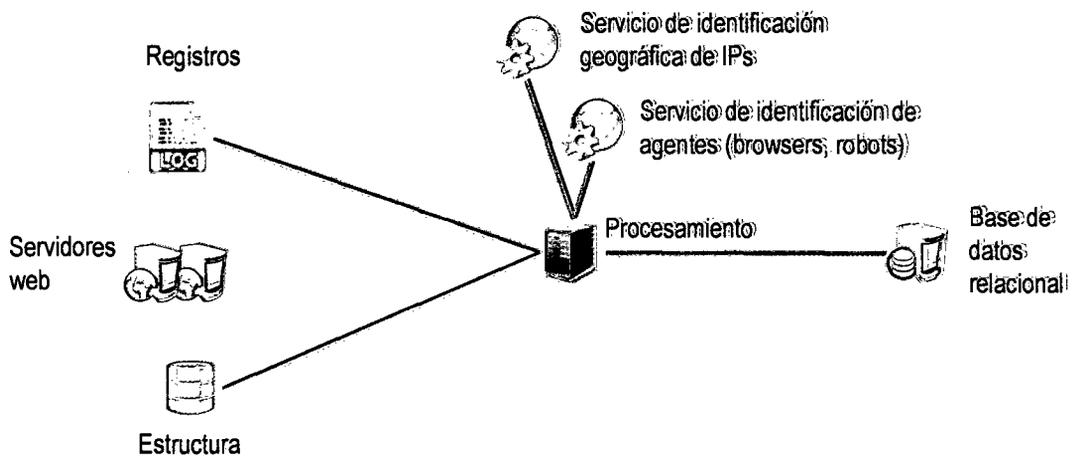
4.3 DISEÑO DEL SISTEMA

4.3.1 COMPONENTE DE DATOS

4.3.1.1 FLUJO DE DATOS

La Figura 24 muestra el flujo general de datos, en esta se muestra que cuatro orígenes, dos internos (estructura de la página web y registros del servidor web) y dos externos a los que se accedería por medio de servicios (bases de datos de IPs y de agentes de navegación).

Figura 24: Flujo de datos



Fuente: (Elaboración propia).

Una aplicación (servidor o PC) –con acceso a los datos de los servidores web y con servicios instalados de identificación geográfica y de agentes de navegación– se encarga de extraer la información de las 4 fuentes, la transforma en una forma normalizada y la carga en la base de datos relacional.

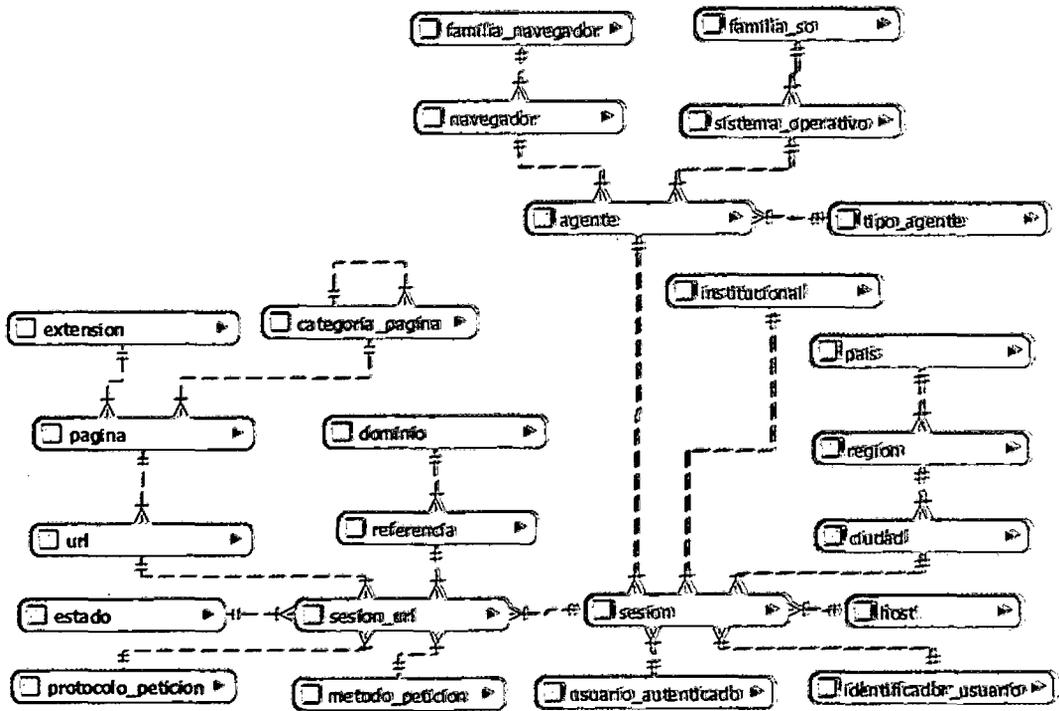
Se debe asegurar que la aplicación (PC o servidor) tenga acceso de lectura en los servidores web (para los registros y la estructura del sitio web), tenga acceso a los servicios de identificación geográfica y de agentes, y además tenga permisos de escritura en la Base de Datos Relacional.

Este procesamiento permite el almacenamiento en una base de datos que da acceso más rápido al SSD.

4.3.1.2 DISEÑO DE LA BASE DE DATOS

La Figura 25 muestra la estructura completa de la Base de Datos, en ella se incorporan las tablas y relaciones establecidas previamente en el análisis de datos.

Figura 25: Estructura completa de la Base de Datos



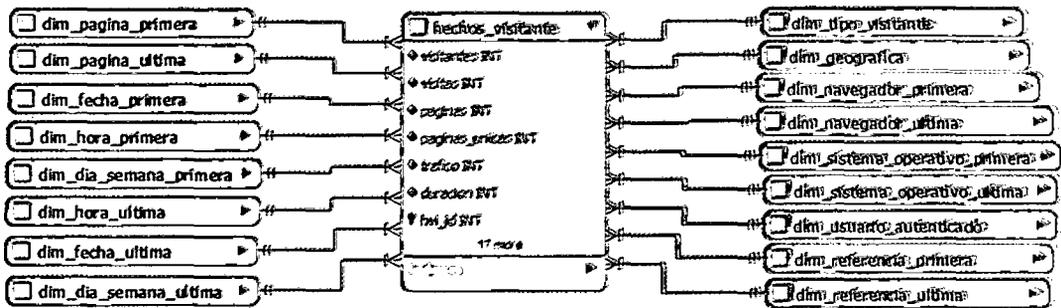
Fuente: (Elaboración propia).

4.3.2 DISEÑO DE LOS MODELOS DE ANÁLISIS

4.3.2.1 MODELO PARA EL ANÁLISIS DE VISITANTES

La Figura 26 describe el diseño lógico para análisis de los visitantes únicos. Esta incorpora una tabla de hechos y 17 dimensiones.

Figura 26: Modelo para análisis de visitantes



Fuente: (Elaboración propia).

La Tabla 6 describe los cinco valores (visitantes, visitas, páginas, páginas únicas, tráfico y duración) de la tabla de hechos para los visitantes.

Tabla 6: Valores para el análisis de visitantes

Descripción (Campo)	Tipo de dato
Identificador (hvi_id)	Entero
Número de visitantes (visitantes)	
Número de visitas (visitas)	Entero
Número de páginas vistas (paginas)	Entero
Número de páginas únicas (paginas_unicas)	Entero
Tráfico de los visitantes medidos en bytes (trafico)	Entero
Duración de los visitantes (duracion)	Entero

Fuente: (Elaboración propia).

La Tabla 7 muestra las dimensiones del modelo para análisis de visitantes

Tabla 7: Dimensiones para el análisis de visitantes

Dimensión (Tabla)	Campos (Tipos de dato)
Geográfica (dim_geografica)	Identificador (Entero)
	Ciudad (VARCHAR)
	Región (VARCHAR)
	País (VARCHAR)
Primer navegador del visitante (dim_navegador_primera)	Identificador (Entero)
	Navegador (VARCHAR)
	Familia del Navegador (VARCHAR)
Último navegador del visitante (dim_navegador_ultima)	Identificador (Entero)
	Navegador (VARCHAR)
	Familia del Navegador (VARCHAR)
Primer sistema operativo del visitante (dim_sistema_operativo_primera)	Identificador (Entero)
	Sistema operativo (VARCHAR)
	Familia de sistema operativo (VARCHAR)
Último sistema operativo del visitante (dim_sistema_operativo_ultima)	Identificador (Entero)
	Sistema operativo (VARCHAR)
	Familia de sistema operativo (VARCHAR)
Nombre del usuario autenticado (dim_usuario_autenticado)	Identificador (Entero)
	usuario (VARCHAR)
Primera referencia del visitante (dim_referencia_primera)	Identificador (Entero)
	referencia (VARCHAR)

Dimensión (Tabla)	Campos (Tipos de dato)
	dominio (VARCHAR)
Última referencia del visitante (dim_referencia_ultima)	Identificador (Entero) referencia (VARCHAR) dominio (VARCHAR)
Día de la semana de la primera visita (dim_dia_semana_primera)	Identificador (Entero) Día de la semana (VARCHAR)
Día de la semana de la última visita (dim_dia_semana_ultima)	Identificador (Entero) Día de la semana (VARCHAR)
Hora de la primera visita (dim_hora_primera)	Identificador (Entero) Segundo (Entero) Minuto (Entero) Hora (Entero)
Hora de la última visita (dim_hora_ultima)	Identificador (Entero) Segundo (Entero) Minuto (Entero) Hora (Entero)
Fecha de la primera visita (dim_fecha_primera)	Identificador (Entero) Número de día del mes (Entero) Mes (Entero) Año (Entero)
Fecha de la última visita (dim_fecha_ultima)	Identificador (Entero) Número de día del mes (Entero)

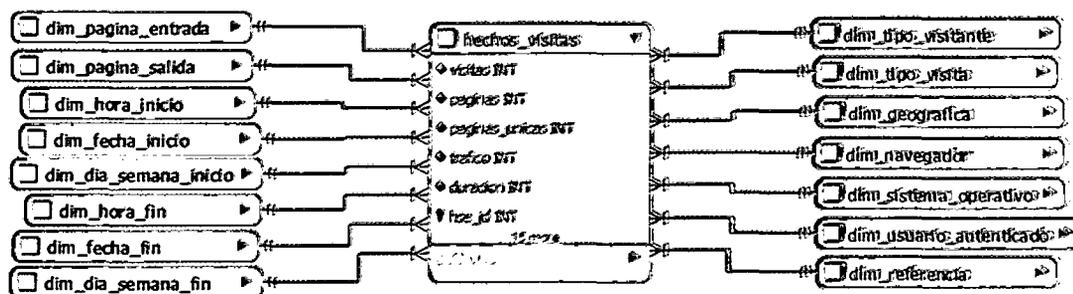
Dimensión (Tabla)	Campos (Tipos de dato)
	Mes (Entero)
	Año (Entero)
	Identificador (Entero)
Primera página vista por el visitante (dim_pagina_primera)	Página (VARCHAR)
	Categoría de nivel 1 (VARCHAR)
	Categoría de nivel 2 (VARCHAR)
	Identificador (Entero)
Última página vista por el visitante (dim_pagina_ultima)	Página (VARCHAR)
	Categoría de nivel 1 (VARCHAR)
	Categoría de nivel 2 (VARCHAR)
	Identificador (Entero)
Tipo de visitante (dim_tipo_visitante)	Familia del visitante (VARCHAR)
	Tipo del visitante (VARCHAR)

Fuente: (Elaboración propia).

4.3.2.2 MODELO PARA EL ANÁLISIS DE VISITAS

La Figura 27 muestra el diseño lógico para análisis de las visitas o sesiones. Esta incorpora una tabla de hechos y 15 dimensiones.

Figura 27: Modelo para análisis de visitas



Fuente: (Elaboración propia).

La Tabla 8 muestra los cinco valores (visitas, páginas, páginas únicas, tráfico y duración) de la tabla de hechos para los visitas.

Tabla 8: Valores para el análisis de visitas

Descripción (Campo)	Tipo de dato
Identificador (hse_id)	Entero
Número de visitas (visitas)	Entero
Número de páginas vistas (paginas)	Entero
Número de páginas únicas (paginas_unicas)	Entero
Tráfico de las visitas medidos en bytes (trafico)	Entero
Duración de las visitas (duracion)	Entero

Fuente: (Elaboración propia).

La Tabla 9 muestra las dimensiones del modelo para análisis de visitas

Tabla 9: Dimensiones para el análisis de visitas

Dimensión (Tabla)	Campos (Tipos de dato)
Geográfica (dim_geografica)	Identificador (Entero)
	Ciudad (VARCHAR)
	Región (VARCHAR)
	País (VARCHAR)
Navegador de la visita (dim_navegador)	Identificador (Entero)
	Navegador (VARCHAR)
	Familia del Navegador (VARCHAR)
Sistema operativo de la visita (dim_sistema_operativo)	Identificador (Entero)
	Sistema operativo (VARCHAR)
	Familia de sistema operativo (VARCHAR)
Nombre del usuario autenticado (dim_usuario_autenticado)	Identificador (Entero)
	usuario (VARCHAR)
Referencia de la visita (dim_referencia)	Identificador (Entero)
	referencia (VARCHAR)
	dominio (VARCHAR)
Día de la semana del inicio de la visita (dim_dia_semana_inicio)	Identificador (Entero)
	Día de la semana (VARCHAR)
Día de la semana del fin de la visita (dim_dia_semana_fin)	Identificador (Entero)
	Día de la semana (VARCHAR)
Hora del inicio de la visita (dim_hora_inicio)	Identificador (Entero)
	Segundo (Entero)

Dimensión (Tabla)	Campos (Tipos de dato)
	Minuto (Entero)
	Hora (Entero)
	Identificador (Entero)
Hora del fin de la visita (dim_hora_fin)	Segundo (Entero)
	Minuto (Entero)
	Hora (Entero)
	Identificador (Entero)
Fecha del inicio de la visita (dim_fecha_inicio)	Número de día del mes (Entero)
	Mes (Entero)
	Año (Entero)
	Identificador (Entero)
Fecha del fin de la visita (dim_fecha_fin)	Número de día del mes (Entero)
	Mes (Entero)
	Año (Entero)
	Identificador (Entero)
Página de entrada (dim_pagina_entrada)	Página (VARCHAR)
	Categoría de nivel 1 (VARCHAR)
	Categoría de nivel 2 (VARCHAR)
	Identificador (Entero)
Página de salida (dim_pagina_salida)	Página (VARCHAR)
	Categoría de nivel 1 (VARCHAR)
	Categoría de nivel 2 (VARCHAR)

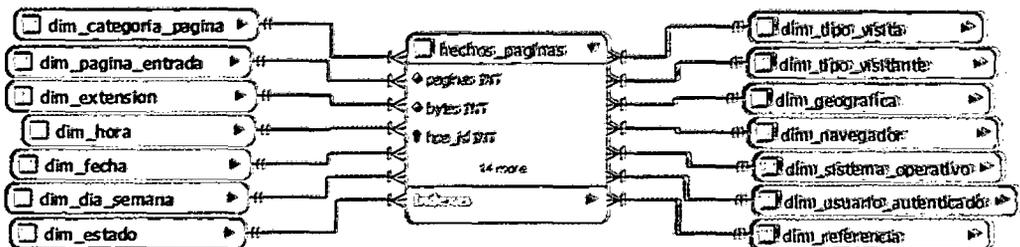
Dimensión (Tabla)	Campos (Tipos de dato)
	Identificador (Entero)
Tipo de la visita (dim_tipo_visita)	Familia del visitante (VARCHAR)
	Tipo del visitante (VARCHAR)
Tipo inicial del visitante (dim_tipo_visitante)	Identificador (Entero)
	Familia del visitante (VARCHAR)
	Tipo del visitante (VARCHAR)

Fuente: (Elaboración propia).

4.3.2.3 MODELO PARA ANÁLISIS DE PÁGINAS

La Figura 28 muestra el diseño lógico para análisis de las páginas, que presenta una tabla de hechos y 14 dimensiones.

Figura 28: Modelo para análisis de páginas



Fuente: (Elaboración propia).

La Tabla 10 muestra los dos valores (páginas y tráfico) de la tabla de hechos para análisis de páginas.

Tabla 10: Valores para el análisis de páginas

Descripción (Campo)	Tipo de dato
Identificador (hpa_id)	Entero
Número de páginas (paginas)	Entero
Tráfico de las visitas medidos en bytes (trafico)	Entero

Fuente: (Elaboración propia).

La Tabla 11 muestra las dimensiones del modelo para análisis de páginas

Tabla 11: Dimensiones para el análisis de páginas

Dimensión (Tabla)	Campos (Tipos de dato)
Geográfica (dim_geografica)	Identificador (Entero)
	Ciudad (VARCHAR)
	Región (VARCHAR)
	País (VARCHAR)
Navegador (dim_navegador)	Identificador (Entero)
	Navegador (VARCHAR)
	Familia del Navegador (VARCHAR)
Sistema operativo (dim_sistema_operativo)	Identificador (Entero)
	Sistema operativo (VARCHAR)
	Familia de sistema operativo (VARCHAR)
Nombre del usuario autenticado (dim_usuario_autenticado)	Identificador (Entero)
	usuario (VARCHAR)

Dimensión (Tabla)	Campos (Tipos de dato)
Referencia (dim_referencia)	Identificador (Entero)
	referencia (VARCHAR)
	dominio (VARCHAR)
Día de la semana (dim_dia_semana)	Identificador (Entero)
	Día de la semana (VARCHAR)
Hora (dim_hora)	Identificador (Entero)
	Segundo (Entero)
	Minuto (Entero)
	Hora (Entero)
Fecha (dim_fecha)	Identificador (Entero)
	Número de día del mes (Entero)
	Mes (Entero)
Estado (dim_estado)	Año (Entero)
	Identificador (Entero)
	Estado (VARCHAR)
Extensión (dim_extension)	Categoría de estado (VARCHAR)
	Identificador (Entero)
	Extensión (VARCHAR)
Página de entrada (dim_pagina_entrada)	Tipo de archivo (VARCHAR)
	Identificador (Entero)
	Página (VARCHAR)
	Categoría de nivel 1 (VARCHAR)

Dimensión (Tabla)	Campos (Tipos de dato)
	Categoría de nivel 2 (VARCHAR)
	Identificador (Entero)
Página	Página (VARCHAR)
(dim_pagina)	Categoría de nivel 1 (VARCHAR)
	Categoría de nivel 2 (VARCHAR)
	Identificador (Entero)
Tipo de la visita	Familia del visitante (VARCHAR)
(dim_tipo_visita)	Tipo del visitante (VARCHAR)
	Identificador (Entero)
Tipo inicial del visitante	Familia del visitante (VARCHAR)
(dim_tipo_visitante)	Tipo del visitante (VARCHAR)

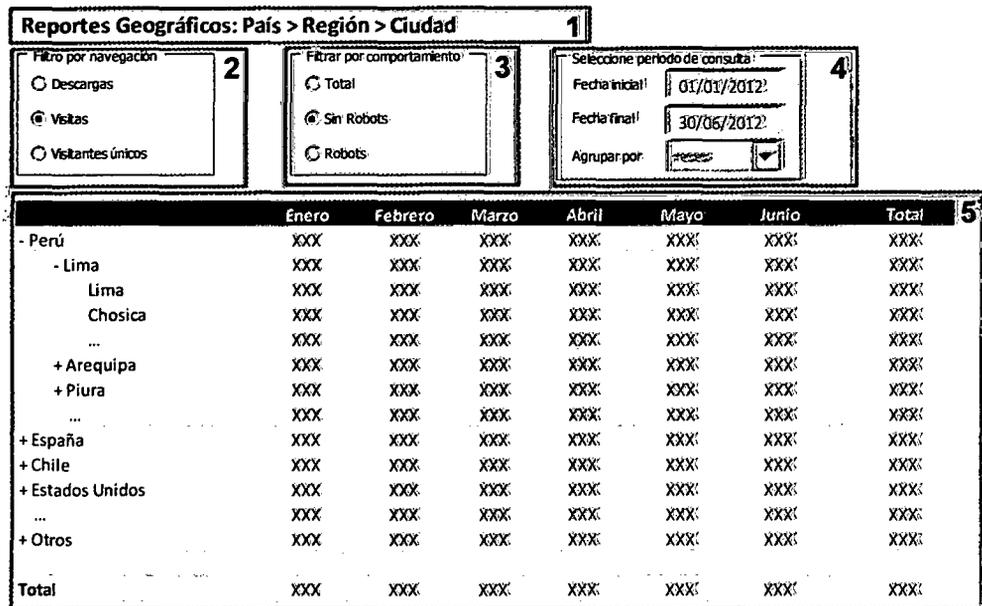
Fuente: (Elaboración propia).

4.3.3 INTERFACES DE USUARIO

Las partes de los reportes del modelo SSD-IWEP se muestran en la Figura 29

- En la parte superior (1) se muestra el título del reporte.
- Debajo del título se muestra el filtro por navegación (2), se refiere a si se requiere mostrar las descargas, visitas o visitantes únicos.
- En la parte central se sitúa el filtro por comportamiento (3), que permite seleccionar si se desea mostrar la actividad de todos los visitantes, de los robots (arañas de la web) o de los no robots (personas).
- En la parte derecha se muestra la parte para seleccionar el periodo de consulta (4), que permite escoger el la fecha inicial, la fecha final y la forma de agrupación del reporte.
- Los cuadros resultantes (5) de la consulta se muestran en la parte inferior.

Figura 29: Partes de los reportes del modelo SSD-IWEP



Fuente: (Elaboración propia).

4.3.3.1 REPORTES GEOGRÁFICOS

Los reportes geográficos muestran información de actividad (visitantes, visitas y descargas) por países y permiten en análisis de profundización por regiones y por ciudades (ver Figura 30).

Figura 30: Diseño de reportes geográficos

Reportes Geográficos: País > Región > Ciudad

Filtro por navegación <input type="radio"/> Descargas <input checked="" type="radio"/> Visitas <input type="radio"/> Visitantes únicos	Filtrar por comportamiento <input type="radio"/> Total <input checked="" type="radio"/> Sin Robots <input type="radio"/> Robots	Seleccione periodo de consulta: Fecha inicial: 01/01/2012 Fecha final: 30/06/2012 Agrupar por: <input type="text"/>
--	---	---

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Total
- Perú	XXX	XXX	XXX	XXX	XXX	XXX	XXX
- Lima	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Lima	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Chosica	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...	---	---	---	---	---	---	---
+ Arequipa	XXX	XXX	XXX	XXX	XXX	XXX	XXX
+ Piura	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...	---	---	---	---	---	---	---
+ España	XXX	XXX	XXX	XXX	XXX	XXX	XXX
+ Chile	XXX	XXX	XXX	XXX	XXX	XXX	XXX
+ Estados Unidos	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...	---	---	---	---	---	---	---
+ Otros	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Total	XXX	XXX	XXX	XXX	XXX	XXX	XXX

Fuente: (Elaboración propia).

Este reporte brinda información importante para analizar el origen geográfico de los visitantes al sitio web.

4.3.3.2 REPORTES DE VISITANTES INSTITUCIONALES

El propósito de los reportes de reportes de visitantes institucionales (ver Figura 31) es evaluar la actividad de usuarios de estas entidades en el sitio web.

Figura 31: Diseño de reportes de visitantes institucionales

Reportes de visitantes institucionales

Filtro por navegación

Descargas

Visitas

Visitantes únicos

Filtrar por comportamiento:

Total

Sin Robots

Robots

Seleccione periodo de consulta:

Fecha inicial: 01/01/2012

Fecha final: 30/06/2012

Agrupar por:

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Total
Institución 1	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Institución 2	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Institución 3	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Institución 4	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Institución 5	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Institución 6	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...
Otros	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Total	XXX	XXX	XXX	XXX	XXX	XXX	XXX

Fuente: (Elaboración propia).

4.3.3.3 REPORTES DE CONTENIDO

Los reportes de contenido (ver Figura 32) permiten visualizar la actividad en el sitio web por contenido, permite profundización por categoría y página.

Figura 32: Diseño de reportes de contenido

Descargas por Contenido: Sección > Categoría > Página

Filtro por navegación

Totales

Máximo 1 por visita

Máximo 1 por visitante

Filtrar por comportamiento:

Total

Sin Robots

Robots

Seleccione periodo de consulta:

Fecha inicial: 01/01/2012

Fecha final: 30/06/2012

Agrupar por:

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Total
- Sección 1	XXX	XXX	XXX	XXX	XXX	XXX	XXX
- Categoría 1	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Página 1	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Página 2	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Página 3	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...
+ Categoría 2	XXX	XXX	XXX	XXX	XXX	XXX	XXX
+ Categoría 3	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...
+ Sección 2	XXX	XXX	XXX	XXX	XXX	XXX	XXX
+ Sección 3	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...
+ Otros	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Total	XXX	XXX	XXX	XXX	XXX	XXX	XXX

Fuente: (Elaboración propia).

4.3.3.4 REPORTES DE PÁGINAS DE ENTRADA

Este reporte muestra cuántas veces las distintas páginas han sido el punto de inicio de una visita o visitante (ver Figura 33).

Este reporte permite agregación por los distintos niveles de categorización del sitio web.

Figura 33: Diseño de reportes de páginas de entrada

Páginas de entrada: Sección > Categoría > Página

Filtro por navegación

Total

De visitas

De visitantes

Filtrar por comportamiento

Total

Sin Robots

Robots

Seleccione período de consulta:

Fecha inicial: 01/01/2012

Fecha final: 30/06/2012

Agrupar por: ▼

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Total
- Sección 1	XXX	XXX	XXX	XXX	XXX	XXX	XXX
- Categoría 1	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Página 1	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Página 2	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Página 3	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...
+ Categoría 2	XXX	XXX	XXX	XXX	XXX	XXX	XXX
+ Categoría 3	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...
+ Sección 2	XXX	XXX	XXX	XXX	XXX	XXX	XXX
+ Sección 3	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...
+ Otros	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Total	XXX	XXX	XXX	XXX	XXX	XXX	XXX

Fuente: (Elaboración propia).

4.3.3.5 REPORTES DE PÁGINAS DE SALIDA

Este reporte muestra cuántas veces las distintas páginas han sido el último punto de una visita o visitante (ver Figura 34).

De manera similar que el reporte anterior, este permite agregación por los distintos niveles de categorización del sitio web.

Figura 34: Diseño de reportes de páginas de salida

Páginas de salida: Sección > Categoría > Página

Filtro por navegación <input checked="" type="radio"/> De visitas <input type="radio"/> De visitantes	Filtrar por comportamiento <input type="radio"/> Total <input checked="" type="radio"/> Sin Robots <input type="radio"/> Robots	Seleccione periodo de consulta: Fecha inicial: 01/01/2012 Fecha final: 30/06/2012 Agrupar por:
--	---	--

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Total
- Sección 1	XXX	XXX	XXX	XXX	XXX	XXX	XXX
- Categoría 1	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Página 1	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Página 2	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Página 3	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...
+ Categoría 2	XXX	XXX	XXX	XXX	XXX	XXX	XXX
+ Categoría 3	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...
+ Sección 2	XXX	XXX	XXX	XXX	XXX	XXX	XXX
+ Sección 3	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...
+ Otros	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Total	XXX	XXX	XXX	XXX	XXX	XXX	XXX

Fuente: (Elaboración propia).

4.3.3.6 REPORTES DE REFERENCIAS

Estos reportes muestran las procedencia de los visitantes al sitio web (ver Figura 35). Las referencias pueden ser agregadas por el dominio al que pertenecen.

En estricto, los reportes de referencias muestran las páginas que los visitantes han visitado previamente.

Figura 35: Diseño de reportes de referencias

Referencias: Dominio > Página

Filtro por navegación <input type="radio"/> De páginas <input checked="" type="radio"/> De visitas <input type="radio"/> De visitantes	Filtrar por comportamiento: <input type="radio"/> Total <input checked="" type="radio"/> Sin Robots <input type="radio"/> Robots	Seleccione periodo de consulta: Fecha inicial: 01/01/2012 Fecha final: 30/06/2012 Agrupar por: <input type="text"/>
--	--	---

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Total
- Sitio externo 1	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Página externa 1	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Página externa 2	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...	XXX	XXX	XXX	XXX	XXX	XXX	XXX
+ Sitio externo 2	XXX	XXX	XXX	XXX	XXX	XXX	XXX
- Sitio externo 3	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Página externa 3	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...	XXX	XXX	XXX	XXX	XXX	XXX	XXX
- Sitio externo 4	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Otros	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Total	XXX	XXX	XXX	XXX	XXX	XXX	XXX

Fuente: (Elaboración propia).

4.3.3.7 REPORTES DE SISTEMAS OPERATIVOS

Los reportes de sistemas operativos muestran la actividad registrada en el sitio web desde los sistemas operativos que utilizan los visitantes al sitio web (ver Figura 36).

Figura 36: Diseño de reportes de sistemas operativos

Reporte por Sistemas Operativos > Versión

Filtro por navegación <input checked="" type="radio"/> Visitas <input type="radio"/> Visitantes	Filtrar por comportamiento: <input checked="" type="radio"/> Total <input type="radio"/> Sin Robots <input type="radio"/> Robots	Seleccione periodo de consulta: Fecha inicial: 01/01/2012 Fecha final: 30/06/2012 Agrupar por: <input type="text"/>
--	--	---

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Total
- Sistema operativo 1	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Versión 1	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Versión 2	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...	XXX	XXX	XXX	XXX	XXX	XXX	XXX
+ Sistema operativo 2	XXX	XXX	XXX	XXX	XXX	XXX	XXX
+ Sistema operativo 3	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Total	XXX	XXX	XXX	XXX	XXX	XXX	XXX

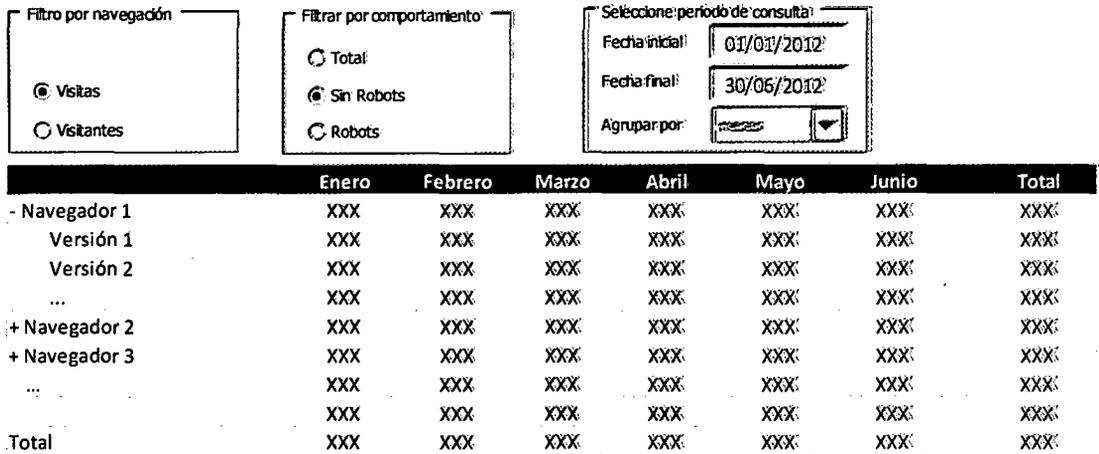
Fuente: (Elaboración propia).

4.3.3.8 REPORTE DE NAVEGADORES

Los reportes de navegadores muestran qué navegadores están utilizando los visitantes al sitio web (ver Figura 37).

Figura 37: Diseño de reportes de navegadores:

Reporte por Navegadores > Versión



Fuente: (Elaboración propia).

4.3.3.9 REPORTE DE ACTIVIDAD EN EL SITIO WEB

Estos reportes muestran la actividad de los visitantes al sitio web durante las distintas horas por día de semana (ver Figura 38).

Figura 38: Diseño de reportes de actividad en el sitio web

Actividad por horas: Sitio web > Página

	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo	Total
00:00	82	65	219	208	87	182	24	867
01:00	106	228	318	247	127	152	85	1,263
02:00	40	148	298	309	155	165	39	1,154
03:00	37	154	315	270	229	207	68	1,280
04:00	151	348	275	313	314	211	27	1,639
05:00	177	261	351	381	310	255	49	1,784
06:00	95	415	367	313	382	259	173	2,004
07:00	389	407	447	456	444	362	157	2,662
08:00	481	492	550	671	392	350	149	3,085
09:00	636	574	702	648	381	430	124	3,495
10:00	706	426	617	671	611	416	208	3,655
11:00	805	629	738	696	589	493	222	4,172
12:00	807	873	842	943	725	525	294	5,009
13:00	813	895	1085	1050	617	681	239	5,380
14:00	1010	965	1032	1177	913	764	379	6,240
15:00	1102	1044	1194	1240	823	835	326	6,564
16:00	1256	1174	1463	1362	945	919	401	7,520
17:00	1251	1379	1427	1596	1135	1026	441	8,255
18:00	1466	1586	1669	1674	1255	1137	556	9,343
19:00	1190	1312	1423	1370	1107	798	367	7,567
20:00	830	931	931	1001	798	668	318	5,477
21:00	627	785	761	657	550	542	199	4,121
22:00	495	618	557	682	556	355	148	3,411
23:00	335	402	584	571	214	368	165	2,639
Total	14887	16111	18165	18506	13,659	12,100	5,158	98,586

Fuente: (Elaboración propia).

4.3.3.10 OTROS REPORTES

El modelo es escalable y permite la obtención de otros reportes adicionales, como el reporte por palabras clave de la página o de los buscadores. El modelo permite también el cálculo de reportes de tiempos de visita, porcentajes de rebotes, entre otros.

Otros reportes como el de errores por página también pueden ser implementados en el sistema.

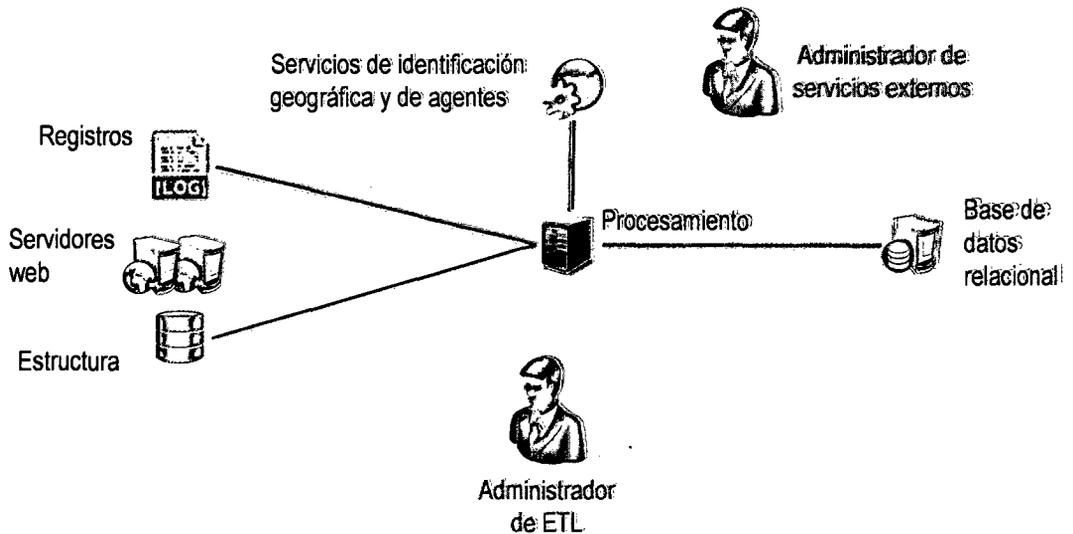
4.3.4 CONSIDERACIONES EN LA ADMINISTRACIÓN Y MANTENIMIENTO DEL SISTEMA

La administración y mantenimiento de un sistema basado en el modelo requerirá la implementación de los siguientes módulos:

- Módulos de gestión de datos
- Módulos de gestión de modelos
- Módulos de gestión de interfaces

Los módulos de gestión de datos (Figura 39) deben permitir al administrador realizar los procesos de extracción, transformación y carga (ETL) internos y administrar los servicios externos de identificación geográfica y de agentes.

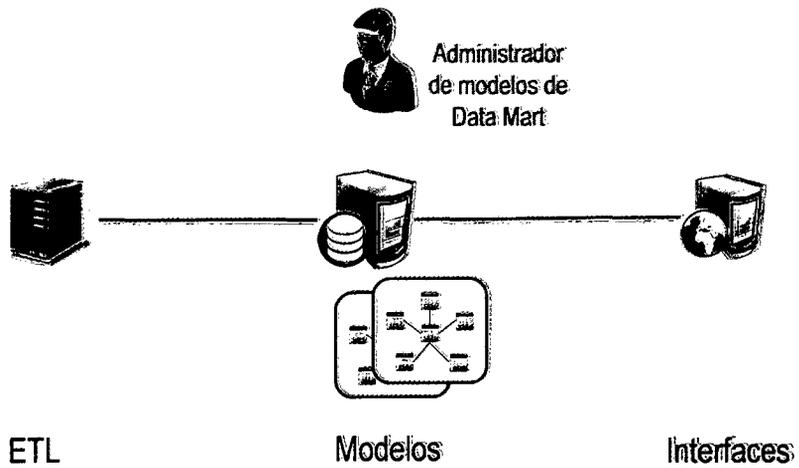
Figura 39: Módulos de gestión de datos



Fuente: (Elaboración propia).

Los módulos de gestión de modelos (Figura 40) deben permitir el manejo de los modelos Data Mart del sistema.

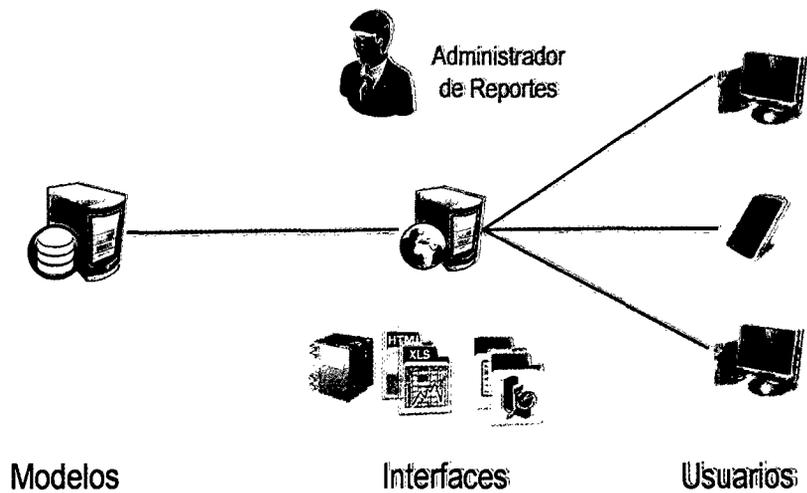
Figura 40: Módulos de gestión de modelos



Fuente: (Elaboración propia).

Los módulos de gestión de interfaces deben permitir controlar los reportes estáticos y dinámicos, así como controlar los accesos de los usuarios a dicho reportes.

Figura 41: Módulos de gestión de interfaces



Fuente: (Elaboración propia).

CAPÍTULO 5

EVALUACIÓN DEL SSD-IWEP

5.1 DESCRIPCIÓN DEL AMBIENTE DEL CASO DE ESTUDIO

Se realizó una encuesta para medir el nivel de aceptación de los reportes por parte de los administradores de sitios web de entidades públicas. Para ello se recolectó los correos y nombres de las personas listadas en el Portal del Estado Peruano – PEP– (<http://www.peru.gob.pe>).

5.2 DISEÑO DE LA VALIDACIÓN DE LOS REPORTES

El autor envió un mensaje a los responsables listados en el PEP con 9 reportes para evaluar su aceptación de los mismos por cada reporte y recoger sus comentarios. Cada reporte permitía una calificación del 1 al 7, donde 1 representaba una opinión desfavorable del reporte (muy poco útil) y 7 representaba la más alta aprobación del mismo (muy útil).

El autor define seis indicadores de aprobación:

- Promedio de aprobación de un reporte: Es el promedio simple de las respuestas de los responsables de contenido de las entidades públicas por cada reporte.
- Grado de aprobación de un reporte: Es la suma del conteo de administradores de sitios web que tienen una calificación de 5, 6 o 7, dividido entre el total de respuestas por cada reporte.

- Promedio de aprobación del modelo: Es el promedio simple de todas las respuestas sobre los reportes del modelo por parte de los responsables de contenido de las entidades públicas.
- Grado de aprobación del modelo: Es la suma del conteo de todas las repuestas de los administradores de sitios web que tienen una calificación mayor a 4 dividido entre el total de respuestas.
- Promedio de aprobación del reporte más valioso: Este indicador toma el valor más alto de aprobación por cada administrador, y luego se hace un promedio simple de estos resultados. En pocas palabras es el promedio de las valoraciones más altas de los responsables de sitios web de entidades públicas.
- Grado de aprobación del reporte más valioso: Es la suma del conteo de administradores de sitios web que tienen una calificación de 5, 6 o 7 en el reporte más valioso, para cada uno de ellos, dividido entre el total responsables de sitios web de entidades públicas.

5.3 RESULTADOS DE APROBACIÓN DE REPORTE

5.3.1 PROMEDIO DE APROBACIÓN DE UN REPORTE

La Tabla 12 muestra el promedio de aprobación por cada reporte, en todos los casos se observa que el promedio es mayor a 5, por lo que se concluye que con este indicador todos los reportes son aprobados.

Tabla 12: Promedio de aprobación de un reporte

Reporte	Promedio de aprobación
Geográfico	5.03
Visitantes institucionales	5.58
Contenido	5.61
Páginas de entrada	5.24
Páginas de salida	5.03

Referencias	5.30
Sistemas Operativos	5.03
Navegadores	5.79
Actividad	6.15

Fuente: (Elaboración propia).

5.3.2 GRADO DE APROBACIÓN DE UN REPORTE

La Tabla 13 muestra el grado de aprobación de los reportes, en la que se observa que en todos los casos, más del 66% de responsables encuentra útil a todos los reportes.

Tabla 13: Grado de aprobación de un reporte

Reporte	Grado de aprobación
Geográfico	69.7%
Visitantes institucionales	72.7%
Contenido	72.7%
Páginas de entrada	69.7%
Páginas de salida	63.6%
Referencias	72.7%
Sistemas Operativos	66.7%
Navegadores	87.9%
Actividad	87.9%

Fuente: (Elaboración propia).

5.3.3 PROMEDIO DE APROBACIÓN DEL MODELO

El promedio de aprobación del modelo es de 5.42, que está por encima del valor medio 4. Lo que indicaría que el modelo es útil para el promedio de responsables del Portal.

5.3.4 GRADO DE APROBACIÓN DEL MODELO

El grado de aprobación del modelo es 73.7%, este valor indicaría que el modelo en conjunto tiene una aprobación adecuada.

5.3.5 PROMEDIO DE APROBACIÓN DEL REPORTE MÁS VALIOSO

Este valor es de 6.76. Este valor está muy cercano al valor máximo, ello indicaría que el modelo presenta en promedio un reporte muy útil para los administradores de sitios web de entidades públicas.

5.3.6 GRADO DE APROBACIÓN DEL REPORTE MÁS VALIOSO

Este valor es de 100%. Ello indicaría que al menos un reporte es aprobado por todos los administradores de sitios web de entidades públicas.

5.4 CARACTERÍSTICAS DEL MODELO

El modelo presenta algunas características importantes que permitirán cumplir las normas que deben cumplir las entidades públicas en cuanto a difusión de información y le permitirá adaptarse a futuros cambios y nuevos requerimientos.

5.4.1 NO INTRUSIVO

El modelo, al utilizar los registros del servidor evita toda interacción directa con los visitantes y no interfiere en su navegación.

5.4.2 SEGURO

El modelo brinda mayor seguridad en el almacenamiento de información debido a que plantea la utilización de Bases de Datos que duplicarían la información de los registros.

5.4.3 MANTIENE LA PRIVACIDAD

Un aspecto fundamental al decidir el tipo de recolección de datos a usar se enfoca en la privacidad de la información de los visitantes.

El modelo SSD-IWEP permite mantener la privacidad de los usuarios, incluso se puede añadir mecanismos adicionales para establecer el anonimato de los visitantes.

Esta característica se debe a que los datos nunca salen de la organización y son manejados por los servidores, que pueden incorporar incluso control de accesos a datos que puedan ser sensibles para los usuarios.

5.4.4 DESEMPEÑO

El modelo presenta un esquema eficiente de manejo de datos, separa el almacenamiento de datos, la consulta de datos y la presentación de los mismos. Esto permite una división de tareas para los servidores.

5.4.5 ESCALABILIDAD

El modelo puede adaptarse a distintas capacidades de procesamiento.

5.4.6 TRANSPARENTE EN DISEÑO

El modelo SSD-IWEP es presentado en esta investigación con todas sus partes y puede ser revisado para añadir mejoras o proponer nuevos cambios de acuerdo a los requerimientos específicos de las entidades públicas.

5.4.7 INDEPENDIENTE DE LA HERRAMIENTA

Un aspecto fundamental del modelo SSD-IWEP es que puede ser implementado en cualquier plataforma que soporte un manejo de BD Relacionales y algún lenguaje de programación.

5.4.8 EXTENSIBLE

El modelo puede incorporar más opciones, según los requerimientos de cada institución. El modelo está preparado para incorporar nuevas funcionalidades como la de Minería de Datos, Modelos Estadísticos, Modelos Matemáticos, etc. de acuerdo a las necesidades de las organizaciones.

5.4.9 EXACTO

El modelo presenta de manera precisa –dada la forma de reconocer visitantes y visitas– toda la actividad de un sitio web.

5.4.10 ACCESIBLE

El modelo puede ser implementado en plataformas web, servidores *stand alone*, computadoras de escritorio. Un servidor web podría a su vez aumentar la accesibilidad de la solución.

5.4.11 FILTRA ROBOTS

El modelo permite el filtro de robots para evaluar las descargas hechas por personas.

5.4.12 MANTIENE LA PROPIEDAD

Tanto los datos como el modelo queda en propiedad de las entidades públicas.

5.4.13 IDENTIFICA TODAS LAS TRANSACCIONES

El hecho que se utilicen los registros del servidor permite que se contabilice todas las transacciones que realiza el servidor.

5.5 APORTES DEL MODELO

El modelo SSD-IWEP mejora la toma de decisiones en la publicación de información en una entidad pública. Esta afirmación fue validada por los administradores de contenido de sitios web de entidades públicas que mostraron una aceptación teniendo en cuenta los seis indicadores propuestos. El modelo SSD-IWEP tiene las características necesarias para su funcionamiento en sitios web de entidades públicas.

El modelo SSD-IWEP mejora la identificación de perfiles de visitantes de una página web en una entidad pública. El modelo permite a los administradores conocer la procedencia geográfica de sus visitantes, las descargas promedio, de dónde vienen, las páginas por donde usualmente entran al sitio y las páginas por donde salen del sitio, los sistemas operativos y navegadores que usan y su actividad en el sitio web. Adicionalmente el sistema permite diferenciar el tráfico de robots (arañas) del tráfico de personas.

El modelo SSD-IWEP mejora la identificación de patrones de descarga de una página web en una entidad pública. El modelo SSD-IWEP permite conocer las páginas más descargadas por categoría, las descargas únicas y reiterativas, la actividad de descargas por horas, las descargas por zonas geográficas.

El modelo SSD-IWEP además permite la escalabilidad para enfrentar futuras necesidades de información de visitantes y descargas de un sitio web público. El modelo SSD-IWEP tiene la base lista para procesos de minería de datos de los registros y análisis de tendencias en los mismos.

CONCLUSIONES Y RECOMENDACIONES

1. CONCLUSIONES

- El aumento de acceso de Internet, aumento de tráfico en las instituciones públicas y las normas peruanas obligan a los administradores de sitios web de entidades públicas a conocer mejor a sus visitantes y sus patrones de descarga con el objetivo de brindar un mejor servicio y evitar el uso de otros canales más costosos.
- Las normas peruanas fomentan la accesibilidad a la información y transparencia de entidades públicas.
- Los modelos basados en Data Mart presentaron una alternativa adecuada para el análisis de uso de un sitio web.
- El modelo SSD-IWEP planteó una extensión de modelos previos y lo orientó al análisis de uso de la página web de entidades públicas. Los reportes del modelo SSD-IWEP permitieron entender mejor a los visitantes y sus descargas en un sitio web de entidad pública.
- El modelo resolvió el problema de mejorar la toma de decisiones en la publicación de información en una entidad pública.
- Los reportes del modelo SSD-IWEP tuvieron aceptación por parte de los administradores de contenido de sitios web de entidades públicas.
- El modelo SSD-IWEP es seguro, mantiene la privacidad de los visitantes, tiene desempeño y escalabilidad a medida, transparente en el diseño, independiente de la herramienta, extensible, exacto, accesible, filtra robots, mantiene la propiedad de los datos y los modelos, identifica todas las transacciones y es no intrusivo.

2. RECOMENDACIONES

- Se requiere más investigaciones sobre penetración de Internet móvil, usuarios proxy y otros mecanismos de acceso a Internet en el país.
- Es necesario estudiar las tecnologías utilizadas en las entidades públicas para evaluar la factibilidad de implementación de distintas herramientas de software.
- Evaluar el desempeño del modelo SSD-IWEP en distintas plataformas, con el objetivo de optimizar su implementación.
- Se recomienda las investigaciones con minería de datos que permitan generar reportes de tendencias, segmentación, entre otros que extenderían el modelo SSD-IWEP.
- Se hace necesario el estudio de más indicadores de éxito de las publicaciones de sitios web de entidades públicas además de los indicadores convencionales (descargas, sesiones, entre otros). En sitios web comerciales, por ejemplo, se estudia mucho distintos ratios de conversión de visitas a compras o valor actual de beneficios.
- Por último, es recomendable el estudio de la aplicación del modelo SSD-IWEP en otras entidades con objetivos de difusión de información o con fines educativos.

GLOSARIO DE TÉRMINOS

ALMACENAMIENTO EN CACHÉ

Aunque hay muchos matices en ella (por ejemplo, almacenamiento en "caché del navegador" y "cache del servidor"), en general se refiere al almacenamiento en caché se refiere a la técnica en la que la mayoría navegadores de Internet guarda una copia de cada página Web, llamada caché, en su memoria.

Muchos proveedores de servicios de Internet y páginas web de grandes organizaciones almacenan en caché para brindar mayor velocidad al acceso de contenidos y reducir el uso de ancho de banda.

CORTAFUEGO (*FIREWALL*)

Un servidor proxy también puede funcionar como un servidor de seguridad en una organización, actuando como un dispositivo intermediario, pero para el propósito de seguridad en lugar de la eficiencia.

Los *firewalls* son utilizados por las organizaciones para proteger a los usuarios internos de las amenazas externas a través de Internet, o para evitar que los empleados tengan acceso a un conjunto específico de sitios web.

DIRECCIÓN IP

Un identificador numérico de una computadora. (El formato de una dirección IP es una dirección numérica de 32 bits escrita como cuatro números separados por puntos, cada número puede ser cero a 255. La dirección IP se puede utilizar para determinar el origen geográfico de un visitante y también se puede utilizar para

determinar la red de computadora en particular de la que los visitantes de un sitio web están viniendo.

DURACIÓN DE LA VISITA

Tiempo medio de permanencia en el lugar (longitud del visitante se queda en el sitio antes de salir). Los sitios que retienen a los visitantes se conocen como sitios "adhesivos".

FLASH

Herramienta de software que permite que las páginas Web incluyan elementos con animación.

JAVA

Lenguaje de programación para aplicaciones en la Web.

IDIOMA DEL VISITANTE

La configuración de idioma en la computadora del visitante.

LOCALIZACIÓN DE UN VISITANTE

La ubicación geográfica del visitante.

NAVEGADOR

El tipo de software navegador que un visitante está utilizando.

ORIENTACIÓN A CAPTURA DE ERRORES

Los registros web se orientaron en un principio a la captura de información técnica (404 errores, tendencias de uso del servidor, tipo de navegador, etc) que no son adecuados para capturar la información comercial.

PÁGINAS VISTAS

El número de veces que una página web especificada ha sido vista, muestra exactamente el contenido que la gente está (o no está) viendo en un sitio web. Cada vez que un visitante pulsa el botón de actualización de la página, se registra otra vista de página.

Este término es referido también como descarga o acceso.

PÁGINAS VISTAS POR VISITANTE

El número de páginas vistas dividido entre el número de visitantes, mide cuántas páginas ven promedio los usuarios cada vez que visita un sitio web.

PÁGINAS DE REFERENCIA / SITIOS (URL)

Indica cómo los visitantes llegan a un sitio web (si escribe la dirección URL directamente en un navegador Web o si hacen clic a través de un enlace de otro sitio).

PALABRAS CLAVE

Si la URL de referencia es un motor de búsqueda, las palabras clave que usa el visitante pueden ser determinados.

PETICIÓN (HIT)

Se refiere a cada elemento de una página Web descargada al navegador Web del visitante.

PORCENTAJE DE REBOTE

El porcentaje de usuarios que abandonan el sitio después de la primera página, que se calcula por el número de visitantes que visitan solamente una sola página dividida por el número total de visitas. El porcentaje de abandonos se utiliza a veces como un indicador de la " adhesividad".

RESOLUCIÓN DE PANTALLA

Los ajustes de pantalla para la computadora del visitante.

RUTAS DE NAVEGACIÓN POR VISITANTES

Cómo navegan los visitantes en el sitio web. Se analiza para este caso las páginas de entrada más comunes (la primera página visitada por el visitante a una página web) y los puntos de salida (la página desde la que un visitante sale de un sitio web).

SERVIDORES PROXY

Un servidor proxy es un servidor de red que actúa como un intermediario entre el equipo del usuario y el servidor web, son usados para mejorar el servicio a grupos de usuarios.

En primer lugar, ahorran los resultados de todas las peticiones de una página Web en particular para una cierta cantidad de tiempo.

SISTEMA OPERATIVO

El sistema operativo específico que el visitante utiliza.

VELOCIDAD DE CONEXIÓN

Tipo de velocidad de acceso que tienen los visitantes a una página web.

VISITA (SESIÓN)

Una visita describe la actividad de un usuario individual en un sitio, es una serie de vistas (accesos) hechas desde una IP teniendo actividad continua con el servidor. La mayoría de las herramientas de análisis pondrá fin a la sesión si el visitante permanece inactivo durante treinta minutos o una hora, aunque ese plazo a menudo se puede ajustar en las opciones de su software.

VISITANTES NUEVOS / VISITANTES RECURRENTE

El número de visitantes por primera vez al sitio en comparación a los visitantes que regresan.

VISITANTE ÚNICO

El número de visitantes a la página web que provienen de direcciones IP únicas.

FUENTE:

(Ferrini & Mohr, 2009)

(Kaushik, Web Analytics: An Hour a Day, 2007)

BIBLIOGRAFÍA

- Apache Software Foundation. (n.d.). Log Files. Retrieved noviembre 15, 2012, from Apache HTTP Server Project: <http://httpd.apache.org/docs/2.0/en/logs.html>
- Barrantes, R. (1999). *Investigación: Un Camino Al Conocimiento. Un Enfoque Cuantitativo Y Cualitativo*. San José, Costa Rica: Editorial Universidad Nacional a Distancia.
- Breslin, M. (2004). Data Warehousing, Battle of the Giants: Comparing the Basics of the Kimball and Inmon Models. *Business Intelligence Journal*, 6-20.
- Chenoweth, T., Schuff, D., & Louis, R. (2003, Diciembre). A method for developing dimensional data marts. *Communications of the ACM - Mobile computing opportunities and challenges*, 46(12), 93-98.
- Cohen, D., & Asín, E. (2000). *Sistemas de información para los negocios: Un enfoque de toma de decisiones*. México: McGraw-Hill.
- Congreso de la República. (2002, Agosto 02). *Ley N° 27806: Ley de Transparencia y Acceso a la Información Pública*. Retrieved Noviembre 24, 2012, from ONGEI: http://www.ongei.gob.pe/banco/ongei_normas_detalle.asp?pk_id_normas=199
- Congreso de la República. (2005, Junio 02). *Ley N° 28530: Ley de Promoción de Acceso a Internet para personas con discapacidad y de adecuación del espacio físico en cabinas públicas de internet*. Retrieved Noviembre 24, 2012, from ONGEI: http://www.ongei.gob.pe/banco/ongei_normas_detalle.asp?pk_id_normas=56

- Congreso de la República. (2007, Setiembre 25). *Ley N° 29091: Ley que modifica el párrafo 38.3 del Artículo 38° de la Ley N° 27444*. Retrieved Noviembre 24, 2012, from ONGEI: http://www.ongei.gob.pe/banco/ongei_normas_detalle.asp?pk_id_normas=91
- Congreso de la República del Perú. (1993). *Constitución de la Republica del Perú*. Retrieved Diciembre 16, 2012, from Congreso de la República del Perú: <http://www2.congreso.gob.pe/sicr/RelatAgenda/constitucion.nsf/NumArticulos/002?openDocument>
- Dragos, S.-M. (2011). Why Google Analytics Cannot Be Used For Educational Web Content. *7th International Conference on Next Generation Web Services Practices* (pp. 113-118). Piscataway: IEEE.
- Escobar, V. H. (2007). *Minería Web de Uso y Perfiles de Usuario: Aplicaciones con Lógica Difusa*. Tesis Doctoral. Granada, España: Universidad de Granada.
- Ferrini, A., & Mohr, J. (2009). Uses, Limitations, and Trends in Web Analytics. In B. Jansen, A. Spink, & I. Taksa, *Handbook of Research on Web Log Analysis*. Hershey, PA: Information Science Reference.
- Hackney, D. (1997). *Understanding and implementing successful data marts*. Reading, Massachusetts: Addison-Wesley Developers Press.
- Hu, X., & Cercone, N. (2004, Julio). A data warehouse/online analytic processing framework for web usage mining and business intelligence reporting. *International Journal of Intelligent Systems - Granular Computing and Data Mining*, 19(7), 585-606.
- Imhoff, C., Gallemmo, N., & Geiger, J. (2003). *Mastering data warehouse design: Relational and dimensional techniques*. Indianapolis: Wiley Pub.
- INEI. (2012, Setiembre 2012). *Informe Técnico N° 03 - Setiembre 2012: Las Tecnologías de Información y Comunicación en los Hogares: Abril-Mayo-Junio, 2012*. Retrieved Noviembre 28, 2012, from INEI: <http://www.inei.gob.pe/web/Biblioinei/BoletinFlotante.asp?file=15083.pdf>

- Inmon, W. (2002). *Building the Data Warehouse* (Tercera Edición ed.). John Wiley & Sons.
- Inmon, W., Strauss, D., & Neushloss, G. (2008). *DW 2.0 the architecture for the next generation of data warehousing*. Amsterdam Boston: Morgan Kaufmann.
- Jansen, B., Taksa, I., & Spink, A. (2009). Research and Methodological Foundations of Transaction Log Analysis. In B. Jansen, I. Taksa, & A. Spink, *Handbook of research on web log analysis*. Hershey, PA: Information Science Reference.
- Joshi, K., Joshi, A., Yesha, Y., & Krishnapuram, R. (1999, Noviembre). Warehousing and Mining Web Logs. *Workshop on Web Information and Data Management. ACM Conference on Information and Knowledge Management (CIKM 99)*.
- Kaushik, A. (2007). *Web Analytics: An Hour a Day*. Wiley Publishing, Inc.
- Kaushik, A. (2010). *Web analytics 2.0 : the art of online accountability & science of customer centricity*. Indianapolis, IN: Wiley.
- Kimball, R. (1998). *The Data Warehouse Lifecycle Toolkit*. New York: Wiley.
- Kimball, R., & Ross, M. (2002). *The Data Warehouse Toolkit, The Complete Guide to Dimensional Modeling* (Segunda Edición ed.). John Wiley & Sons.
- King, A. (2008). *Website optimization*. Farnham: O'Reilly.
- Kumar, K., & Reddy, R. (2006). Role of metadata in the datawarehousing environment. Master Thesis. Luleå University of Technology.
- Li, R. (2000). Web clickstream data analysis using a dimensional data warehouse. *Tesis para obtener el grado de Master of Engineering in Computer Science and Engineering, MIT*.
- Mattison, R. (2006). *The data warehousing handbook*. Oakwood Hills: XIT Press.

- Nader, J. (2004). Sistema de Apoyo Gerencial Universitario. Tesis de Magister en Ingeniería de Software. Instituto Tecnológico de Buenos Aires.
- OSIPTEL. (2012). *Informe de Gestión 2007-2012*. Retrieved Noviembre 18, 2012, from [Memorias Anuales: http://www.osiptel.gob.pe/WebSiteAjax/WebFormGeneral/Publicaciones/wfrm_Consulta_Informacion_publicaciones.aspx?CodInfo=0&CodiCat=3&CodiSubcat=568&TituloInformacion=Memorias%20Anuales](http://www.osiptel.gob.pe/WebSiteAjax/WebFormGeneral/Publicaciones/wfrm_Consulta_Informacion_publicaciones.aspx?CodInfo=0&CodiCat=3&CodiSubcat=568&TituloInformacion=Memorias%20Anuales)
- Peterson, E. (2004). *Web analytics demystified : a marketer's guide to understanding how your web site affects your business*. Portland, OR San Leandro: Celilo Group Media CafePress.
- Power, D. (2008a). Decision Support Systems Concept. In *Encyclopedia of Decision Making and Decision Support Technologies* (Adam, Frédéric; Humphreys, Patrick ed.). Hershey: Information Science Reference.
- Power, D. (2008b). Decision Support Systems: A Historical Overview. In F. Burstein, & C. Holsapple (Eds.), *Handbook on Decision Support Systems 1* (p. 854).
- Power, D. (n.d.). *Decision Support System*. Retrieved Noviembre 15, 2012, from DSSResources.COM: <http://dssresources.com/glossary/48.php>
- Power, D. J., Burstein, F., & Sharda, R. (2011). Reflections on the Past and Future of Decision Support Systems: Perspective of Eleven Pioneers. In *Decision Support: An Examination of the DSS Discipline* (Sharda, Ramesh; Voß, Stefan ed., p. 25). New York: Springer.
- Presidencia del Consejo de Ministros. (2001, Mayo 23). *Decreto Supremo N° 060-2001-PCM*. Retrieved Noviembre 24, 2012, from ONGEI: http://www.ongei.gob.pe/banco/ongei_normas_detalle.asp?pk_id_normas=12
- Presidencia del Consejo de Ministros. (2004, Noviembre 08). *Decreto Supremo N° 059-2004-PCM*. Retrieved Noviembre 24, 2012, from ONGEI: http://www.ongei.gob.pe/banco/ongei_normas_detalle.asp?pk_id_normas=51

- Presidencia del Consejo de Ministros. (2006, Junio 21). *Decreto Supremo N° 032-2006-PCM*. Retrieved Noviembre 24, 2012, from ONGEI: http://www.ongei.gob.pe/banco/ongei_normas_detalle.asp?pk_id_normas=71
- Presidencia del Consejo de Ministros. (2008, Enero 17). *Decreto Supremo N° 004-2008-PCM*. Retrieved Noviembre 24, 2012, from ONGEI: http://www.ongei.gob.pe/banco/ongei_normas_detalle.asp?pk_id_normas=95
- Presidencia del Consejo de Ministros. (2009, Marzo 25). *Resolución Ministerial N° 126-2009-PCM*. Retrieved 11 24, 2012, from ONGEI: http://www.ongei.gob.pe/banco/ongei_normas_detalle.asp?pk_id_normas=125
- Presidencia del Consejo de Ministros. (2010, Junio 03). *Decreto Supremo N° 063-2010-PCM*. Retrieved Noviembre 24, 2012, from ONGEI: http://www.ongei.gob.pe/banco/ongei_normas_detalle.asp?pk_id_normas=150
- Presidencia del Consejo de Ministros. (2010, Diciembre 14). *Resolucion Ministerial N° 374-2010-PCM*. Retrieved 11 24, 2012, from ONGEI: http://www.ongei.gob.pe/banco/ongei_normas_detalle.asp?pk_id_normas=181
- Ravikumar, G., Manjunath, T., Ravindra, S., & Archana, R. (2011, Abril). A study on design and analysis of web mart mining and its relevance today. *International Journal of Engineering Science and Technology*, 3(4), 3141-3152.
- Sauter, V. (2010). *Decision support systems for business intelligence*. Hoboken, N.J: Wiley.
- Serra, S., Campos, L., Machado, M., & Pires, P. (2003). A data mart approach for monitoring Web services usage and evaluating quality of services. *XVIII Simpósio Brasileiro de Banco de Dados*. Brasil.
- Teixeira, C., & David, G. (2005). A Data Webhouse to monitor the use of a Web Based Higher Education Information System. *EUNIS 2005 - Leadership and Strategy in a Cyber-Infrastructure World*. Reino Unido.

Turban, E., & Volonino, L. (2011). *Information technology for management : improving strategic and operational performance*. Hoboken, NJ: John Wiley.

Velásquez, J., & Palade, V. (2008). *Adaptive web sites a knowledge extraction from web data approach*. Amsterdam, Países Bajos: IOS Press.

ANEXOS

1. ENCUESTA ENVIADA A LOS RESPONSABLES DE CONTENIDO DE SITIOS WEB DE ENTIDADES PÚBLICAS.

Estimado Sr [Nombres] / Responsable del Portal de [Sitio web]:

Estoy llevando a cabo una investigación sobre los reportes de uso de páginas web de entidades públicas. Como resultado de la investigación planteo 9 reportes clave que apoyarán la toma de decisiones en la difusión de información en los sitios web de entidades públicas.

En este correo le adjunto la encuesta y espero que pueda responderme sobre la utilidad de estos reportes para la toma de decisiones. Puede usar cualquiera de los 2 archivos.

Solo tiene que calificar en la escala de 1 (poco útil) al 7 (muy útil) cada uno de los reportes, en la parte superior con fondo amarillo de cada hoja y si desea añada un comentario.

La investigación estará disponible en la Sección de Postgrado de la Facultad de Ingeniería Industrial y de Sistemas de la Universidad Nacional de Ingeniería.

Si tiene cualquier consulta, por favor, envíeme un correo.

Agradezco de antemano su respuesta.

Atentamente,

Ronald Cuela

Nota:

1.- Todas las respuestas serán tratadas de manera anónima.

2.- La información es obtenida de registros de servidores ficticios.

3.- Visitante, visita y descarga

- Visitante se considera a una dirección IP que ve la página (por ejemplo: Una PC)

- Visita es una sesión en la página hecha por un visitante.

- Descarga es el requerimiento de un archivo de la página web.

4.- Total, sin robots y solo robots.

- Los robots en general son aplicaciones que visitan una página web para almacenar información (pueden ser buscadores, directorios). La característica principal es que hacen bastantes visitas o descargas a la página y podrían distorsionar la información de visitas, visitantes y/o descargas.

- La opción "Sin robots" filtra los robots que han sido identificados por el sistema.

Archivos adjuntos Encuesta.xls y Encuesta.xlsx

2. PANTALLAS DE LA ENCUESTA ENVIADA A LOS RESPONSABLES DE SITIOS WEB DE ENTIDADES PÚBLICAS

2.1. PANTALLA DEL REPORTE REGIONAL (GEOGRÁFICO)

¿Cuál es su opinión sobre este reporte?

Poco útil 1 2 3 4 5 6 7 Muy útil

Comentario...

Reportes Regionales: País > Región > Ciudad

Filtro por navegación

Descargas

Visitas

Visitantes únicos

Filtrar por comportamiento

Total

Sin Robots

Robots

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Total
- Perú	XXX	XXX	XXX	XXX	XXX	XXX	XXX
- Lima	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Lima	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Chosica	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...	XXX	XXX	XXX	XXX	XXX	XXX	XXX
+ Arequipa	XXX	XXX	XXX	XXX	XXX	XXX	XXX
+ Piura	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...	XXX	XXX	XXX	XXX	XXX	XXX	XXX
+ España	XXX	XXX	XXX	XXX	XXX	XXX	XXX
+ Chile	XXX	XXX	XXX	XXX	XXX	XXX	XXX
+ Estados Unidos	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...	XXX	XXX	XXX	XXX	XXX	XXX	XXX
+ Otros	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Total	XXX	XXX	XXX	XXX	XXX	XXX	XXX

2.2. PANTALLA DEL REPORTE DE VISITANTES INSTITUCIONALES

¿Cuál es su opinión sobre este reporte?

Poco útil 1 2 3 4 5 6 7 Muy útil

Comentario...

Reportes de visitantes institucionales

Filtro por navegación

Descargas

Visitas

Visitantes únicos

Filtrar por comportamiento

Total

Sin Robots

Robots

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Total
Institución 1	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Institución 2	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Institución 3	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Institución 4	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Institución 5	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Institución 6	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...
Otros	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Total	XXX	XXX	XXX	XXX	XXX	XXX	XXX

2.3. PANTALLA DEL REPORTE DE CONTENIDO

¿Cuál es su opinión sobre este reporte?

Poco útil 1 2 3 4 5 6 7 Muy útil

Comentario...

Descargas por Contenido: Sección > Categoría > Página

Filtro por navegación

Totales

Máximo 1 por visita

Máximo 1 por visitante

Filtrar por comportamiento

Total

Sin Robots

Robots

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Total
- Sección 1	XXX	XXX	XXX	XXX	XXX	XXX	XXX
- Categoría 1	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Página 1	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Página 2	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Página 3	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...
+ Categoría 2	XXX	XXX	XXX	XXX	XXX	XXX	XXX
+ Categoría 3	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...
+ Sección 2	XXX	XXX	XXX	XXX	XXX	XXX	XXX
+ Sección 3	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...
+ Otros	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Total	XXX	XXX	XXX	XXX	XXX	XXX	XXX

2.4. PANTALLA DEL REPORTE DE PÁGINAS DE ENTRADA

¿Cuál es su opinión sobre este reporte?

Poco útil 1 2 3 4 5 6 7 Muy útil

Comentario...

Páginas de entrada: Sección > Categoría > Página

Filtro por navegación

Totales

De visitas

De visitantes

Filtrar por comportamiento

Total

Sin Robots

Robots

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Total
- Sección 1	XXX	XXX	XXX	XXX	XXX	XXX	XXX
- Categoría 1	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Página 1	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Página 2	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Página 3	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...
+ Categoría 2	XXX	XXX	XXX	XXX	XXX	XXX	XXX
+ Categoría 3	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...
+ Sección 2	XXX	XXX	XXX	XXX	XXX	XXX	XXX
+ Sección 3	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...
+ Otros	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Total	XXX	XXX	XXX	XXX	XXX	XXX	XXX

2.5. PANTALLA DEL REPORTE DE PÁGINAS DE SALIDA

¿Cuál es su opinión sobre este reporte?

Poco útil 1 2 3 4 5 6 7 Muy útil

Comentario...

Páginas de salida: Sección > Categoría > Página

Filtro por navegación

Totales

De visitas

De visitantes

Filtrar por comportamiento

Total

Sin Robots

Robots

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Total
- Sección 1	XXX	XXX	XXX	XXX	XXX	XXX	XXX
- Categoría 1	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Página 1	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Página 2	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Página 3	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...
+ Categoría 2	XXX	XXX	XXX	XXX	XXX	XXX	XXX
+ Categoría 3	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...
+ Sección 2	XXX	XXX	XXX	XXX	XXX	XXX	XXX
+ Sección 3	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...
+ Otros	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Total	XXX	XXX	XXX	XXX	XXX	XXX	XXX

2.6. PANTALLA DEL REPORTE DE REFERENCIAS

¿Cuál es su opinión sobre este reporte?

Poco útil 1 2 3 4 5 6 7 Muy útil

Comentario...

Referencias: Dominio > Página

Filtro por navegación

De páginas

De visitas

De visitantes

Filtrar por comportamiento

Total

Sin Robots

Robots

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Total
- Sitio externo 1	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Página externa 1	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Página externa 2	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...	XXX	XXX	XXX	XXX	XXX	XXX	XXX
+ Sitio externo 2	XXX	XXX	XXX	XXX	XXX	XXX	XXX
- Sitio externo 3	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Página externa 3	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...	XXX	XXX	XXX	XXX	XXX	XXX	XXX
- Sitio externo 4	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Otros	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Total	XXX	XXX	XXX	XXX	XXX	XXX	XXX

2.7. PANTALLA DEL REPORTE DE SISTEMAS OPERATIVOS

¿Cuál es su opinión sobre este reporte?

Poco útil 1 2 3 4 5 6 7 Muy útil

Comentario...

Reporte por Sistemas Operativos > Versión

Filtro por navegación

Descargas

Visitas

Visitantes

Filtrar por comportamiento

Total

Sin Robots

Robots

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Total
- Sistema operativo 1	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Versión 1	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Versión 2	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...	XXX	XXX	XXX	XXX	XXX	XXX	XXX
+ Sistema operativo 2	XXX	XXX	XXX	XXX	XXX	XXX	XXX
+ Sistema operativo 3	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Total	XXX	XXX	XXX	XXX	XXX	XXX	XXX

2.8. PANTALLA DEL REPORTE DE NAVEGADORES

¿Cuál es su opinión sobre este reporte?

Poco útil 1 2 3 4 5 6 7 Muy útil

Comentario...

Reporte por Navegadores > Versión

Filtro por navegación

Descargas

Visitas

Visitantes

Filtrar por comportamiento

Total

Sin Robots

Robots

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Total
- Navegador 1	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Versión 1	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Versión 2	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...	XXX	XXX	XXX	XXX	XXX	XXX	XXX
+ Navegador 2	XXX	XXX	XXX	XXX	XXX	XXX	XXX
+ Navegador 3	XXX	XXX	XXX	XXX	XXX	XXX	XXX
...	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Total	XXX	XXX	XXX	XXX	XXX	XXX	XXX

2.9. PANTALLA DEL REPORTE DE ACTIVIDAD POR HORAS

¿Cuál es su opinión sobre este reporte?

Poco útil 1 2 3 4 5 6 7 Muy útil

Comentario...

Actividad por horas: Sitio web > Página

Filtro por navegación

- De descargas
 De visitas
 De visitantes

Filtrar por comportamiento

- Total
 Sin Robots
 Robots

	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo	Total
00:00	82	65	219	208	87	182	24	867
01:00	106	228	318	247	127	152	85	1,263
02:00	40	148	298	309	155	165	39	1,154
03:00	37	154	315	270	229	207	68	1,280
04:00	151	348	275	313	314	211	27	1,639
05:00	177	261	351	381	310	255	49	1,784
06:00	95	415	367	313	382	259	173	2,004
07:00	389	407	447	456	444	362	157	2,662
08:00	481	492	550	671	392	350	149	3,085
09:00	636	574	702	648	381	430	124	3,495
10:00	706	426	617	671	611	416	208	3,655
11:00	805	629	738	696	589	493	222	4,172
12:00	807	873	842	943	725	525	294	5,009
13:00	813	895	1085	1050	617	681	239	5,380
14:00	1010	965	1032	1177	913	764	379	6,240
15:00	1102	1044	1194	1240	823	835	326	6,564
16:00	1256	1174	1463	1362	945	919	401	7,520
17:00	1251	1379	1427	1596	1135	1026	441	8,255
18:00	1466	1586	1669	1674	1255	1137	556	9,343
19:00	1190	1312	1423	1370	1107	798	367	7,567
20:00	830	931	931	1001	798	668	318	5,477
21:00	627	785	761	657	550	542	199	4,121
22:00	495	618	557	682	556	355	148	3,411
23:00	335	402	584	571	214	368	165	2,639
Total	14887	16111	18165	18506	13,659	12,100	5,158	98,586

3. RESPUESTAS DE LOS ADMINISTRADORES DE CONTENIDO DE SITIOS WEB DE ENTIDADES PÚBLICAS

Institución	Calificación de los reportes									Máx.
	1	2	3	4	5	6	7	8	9	
institucion 1	7	7	7	4	4	7	5	5	6	7
institucion 2	6	4	6	6	6	5	5	6	4	6
institucion 3	7	7	7	6	4	4	6	7	7	7
institucion 4	6	6	6	4	5	5	4	6	6	6
institucion 5	5	5	5	5	5	5	6	6	6	6
institucion 6	7	6	7	5	5	7	5	6	7	7
institucion 7	5	3	7	4	4	4	5	4	1	7
institucion 8	7	7	7	6	5	5	4	5	7	7
institucion 9	7	7	7	7	7	7	7	7	7	7
institucion 10	5	6	6	6	6	7	7	7	7	7
institucion 11	1	3	6	6	1	2	4	5	7	7
institucion 12	3	5	3	3	4	3	3	3	5	5
institucion 13	4	7	6	7	4	5	7	7	7	7
institucion 14	5	7	6	6	7	6	5	7	7	7
institucion 15	5	2	2	3	3	6	3	5	6	6
institucion 16	1	7	3	4	4	5	4	2	7	7
institucion 17	6	6	6	6	6	6	6	6	7	7
institucion 18	5	7	7	1	1	4	4	4	5	7
institucion 19	5	6	4	5	5	6	6	6	6	6
institucion 20	7	7	7	7	7	7	7	7	7	7
institucion 21	7	7	7	7	7	6	5	5	4	7
institucion 22	3	4	4	4	4	4	6	7	7	7
institucion 23	4	5	3	4	4	5	6	7	6	7
institucion 24	5	4	4	5	5	2	5	6	7	7
institucion 25	4	4	3	3	4	4	3	6	7	7
institucion 26	3	5	6	6	5	6	4	6	5	6
institucion 27	6	6	5	5	5	6	6	6	7	7
institucion 28	7	7	7	7	7	7	6	6	7	7
institucion 29	6	4	4	7	7	6	3	7	3	7
institucion 30	5	7	7	6	7	7	7	7	7	7
institucion 31	7	7	7	6	6	7	6	6	7	7
institucion 32	2	2	6	5	5	7	5	5	7	7
institucion 33	3	7	7	7	7	2	1	6	7	7
Promedio	5.03	5.58	5.61	5.24	5.03	5.30	5.03	5.79	6.15	6.76

Alfa de Cronbach: 0.82