

Universidad de las Ciencias Informáticas

Facultad 1



**Módulo de procesamiento estadístico para el
apoyo a la toma de decisiones del motor de
búsqueda Orión**

**Trabajo de Diploma para optar por el Título de
Ingeniero en Ciencias Informáticas**

Autora:

Aimet Cabrera Martínez

Tutores:

MSc. Aneyty Martín García

Ing. Leiny Amel Pons Flores

Ing. Iguelit B. Morán Hernández

La Habana, junio del 2017

“Año 58 del triunfo de la Revolución”



“La única lucha que se pierde es la que se abandona”

Ernesto Guevara de la Serna

Declaración de Autoría

Declaro ser la autora del presente trabajo de diploma y reconozco a la Universidad de las Ciencias Informáticas los derechos patrimoniales del mismo, con carácter exclusivo. Para que así conste firmo la presente a los ____ días del mes de _____ del año _____.

Firma de la autora

Aimet Cabrera Martínez

Firma de la tutora

MSc. Aneyty Martín García

Firma del tutor

Ing. Leiny Amel Pons Flores

Firma de la tutora

Ing. Iguelit B. Morán Hernández

Dedicatoria

A lo largo de mi vida han existido personas que me han brindado su amor incondicional para poder enfrentarme a todos los obstáculos que se me han impuesto en la vida. Este éxito se lo dedico a mi mamá que desde donde está se siente orgullosa de ver que por fin pude lograr su sueño.

Agradecimientos

No me alcanzarían ni las palabras ni las páginas para agradecer a quienes de una forma u otra han contribuido para que este día no haya sido una fantasía y poder ser quien soy.

A mi mamá y mi papá que me han guiado en la vida y me han acompañado en el largo camino de mi educación, por sus deseos de verme como una profesional y contar con la seguridad, apoyo y confianza de ambos por siempre estar orgullosos de mí. A ustedes que la vida sería demasiada corta para agradecerles y pagarles todo lo que han hecho por mí.

A mi hermana que siempre ha sido mi ejemplo a seguir y que me enseñó a estar segura de mi misma; a mi hermano Maury Yuset y a mis sobrinas Nayla, Maura Carmen y Verónica Andrea que son mis rayitos de sol.

A mi familia por su apoyo y amor incondicional sobre todo a mi tía Lety por ser como una segunda madre para mí.

A mi novio y amigo que me ha estado acompañando durante estos 5 años y que me ha aguantado todas mis malcriadeces pero sobre todo me ayudado en las buenas y en las malas.

A mis tutores Aneyty y Leiny por toda la ayuda que me han brindado en este trayecto tan difícil.

Al profesor Paul que cada vez que lo necesitaba estaba dispuesto ayudarme sin ser mi tutor.

A mi amiga y tutora Iguelit que ha sido un gran apoyo durante todo mi estudio.

A mi grupo que compartimos juntos estos 5 años en especial a Daynis que a pesar que no nos soportábamos cuando nos conocimos terminó siendo una gran compañera y amiga en todo momento.

A todas mis amistades más cercana por los buenos momentos que hemos compartido todo este tiempo Betsy, Raisa, Lisbeth, Carlos y Yosel.

A todos aquellos que de una forma u otra me han brindado su apoyo para ser hoy una profesional.

Gracias

Resumen

En el presente trabajo se desarrolló un módulo de procesamiento estadístico para el apoyo a la toma de decisiones para el motor de búsqueda Orión, con el objetivo de agrupar la información de las búsquedas de los usuarios teniendo en cuenta las categorías del contenido de los sitios. El módulo desarrollado posee características y funcionalidades que posibilitan a los usuarios administradores tomar decisiones a partir del análisis estadístico de las búsquedas. Se realizó una investigación de sistemas relacionados con el análisis estadístico de la información para tener en cuenta los elementos a utilizar en el desarrollo del módulo. Para la implementación de la propuesta de solución se empleó la metodología AUP-UCI y la siguiente base tecnológica: el marco de trabajo *Symfony 2.8*, *Bootstrap* en su versión 3.0 para el diseño de la interfaz web, como sistema gestor de bases de datos *PostgreSQL* versión 9.2 y *Solar* versión 4.10.3, *HighChartsJS* como librería para generar los gráficos y *Visual Paradigm 8.0* como herramienta para el modelado. Para darle validez a la propuesta de solución se definió una estrategia de prueba que permitió asegurar que el módulo es funcional y cumple con los requisitos identificados con el cliente. Se realizaron pruebas funcionales, de integración, de seguridad y de usabilidad, de esta manera se garantizó que el módulo tuviera la menor cantidad de errores posibles.

Palabras clave: análisis estadístico, buscador Orión, categoría, recuperación de información, toma de decisiones.

Índice

Introducción	11
Capítulo 1: Fundamentación teórica del módulo procesamiento estadístico para el apoyo a la toma de decisiones en el motor de búsqueda Orión.	17
1.1 Introducción.....	17
1.2 Marco conceptual	17
1.3 Estudio de sistemas homólogos en el ámbito internacional	20
1.4 Estudios de sistemas homólogos en el ámbito nacional	22
1.5 Selección del entorno de desarrollo para la construcción de la solución.....	23
1.6 Conclusiones del capítulo	30
Capítulo 2: Diseño del módulo procesamiento estadístico para el apoyo a la toma de decisiones para el motor de búsqueda Orión	1
2.1 Introducción.....	1
2.2 Modelo del dominio	1
2.3 Especificación de requisitos de <i>software</i>	2
2.4 Historias de usuario.....	4
2.5 Estilo arquitectónico	8
2.6 Patrones de diseño.....	9
2.7 Modelo de Diseño.....	11
2.8 Modelo de datos.....	12
2.9 Modelo de despliegue	13
2.10 Conclusiones del capítulo	14
Capítulo 3: Implementación y validación del módulo procesamiento estadístico para el apoyo a la toma de decisiones para el motor de búsqueda Orión	15
3.1 Introducción.....	15
3.2 Modelo de implementación.....	15

3.3 Diagrama de componentes.....	15
3.4 Estándares de codificación utilizados	16
3.5 Validación del módulo	17
3.6 Criterio de expertos	28
3.7 Conclusiones del capítulo.....	30
Conclusiones Generales.....	31
Recomendaciones	32
Referencias Bibliográficas.....	33
Anexos.....	40
Anexo # 1: Encuestas.....	40
Anexo # 2: Respuestas dadas por los expertos para cada indicador.....	46
Anexo # 3: Diagrama de componente	46
Anexo # 4: Diagrama de diseño de clases.....	48

Índice de figuras

Figura 1. Estadística del portal web CUBANIC	22
Figura 2. Modelo de dominio.....	2
Figura 3. Arquitectura del sistema.....	9
Figura 4. Diagrama de clases del diseño de publicaciones por categoría	12
Figura 5. Modelo de datos	13
Figura 6. Diagrama de despliegue	14
Figura 7. Diagrama de componente de publicaciones por categorías	16
Figura 8. StatisticBundle.php	17
Figura 9. StatisticBundle.php	17
Figura 10. StatisticBundle.php	17
Figura 11. Prueba de seguridad iteración 1.....	24
Figura 12. Prueba de seguridad iteración 2.....	25
Figura 13. Nivel de usabilidad después de corregidos indicadores de usabilidad.....	27
Figura 14. Distribución de frecuencia múltiple de la prueba de actitud	30
Figura 15. Comportamiento de las categorías.....	46
Figura 16. Categoría con menos publicaciones	47
Figura 17. Publicaciones por categorías	47
Figura 18. Comportamiento de las categorías.....	48
Figura 19. Categoría con menos publicaciones	48
Figura 20. Publicaciones por categorías	49

Índice de tablas

Tabla 1. Operacionalización de la variable dependiente	14
Tabla 2. Operacionalización de la variable independiente.....	14
Tabla 3. Requisitos funcionales y no funcionales	2
Tabla 4. Historia de usuario #1	5
Tabla 5. Historia de usuario #2	6
Tabla 6. Historia de usuario #10	6
Tabla 7. Historia de usuario #12	7
Tabla 8. Caso de prueba # 1.....	18
Tabla 9. Variables empleadas en el diseño del caso de prueba # 1	19
Tabla 10. Caso de prueba # 2.....	20
Tabla 11. Variables empleadas en el diseño del caso de prueba # 2.....	21
Tabla 12. Caso de prueba # 3.....	21
Tabla 13. Variables empleadas en el diseño del caso de prueba # 3.....	22
Tabla 14. Cantidad de no conformidades por cada iteración las pruebas	22
Tabla 15. Resultados de prueba de usabilidad utilizando lista de chequeo	25
Tabla 16. Cantidad de errores por cada iteración las pruebas de integración	28
Tabla 17. Clasificación	28
Tabla 18. Valoración de expertos en cuanto al sistema	29
Tabla 19. Cuestionario de actitudes realizado al encuestado # 1.....	40
Tabla 20. Cuestionario de actitudes realizado al encuestado # 2.....	41
Tabla 21. Cuestionario de actitudes realizado al encuestado # 3.....	41
Tabla 22. Cuestionario de actitudes realizado al encuestado # 4.....	42
Tabla 23. Cuestionario de actitudes realizado al encuestado # 5.....	42
Tabla 24. Cuestionario de actitudes realizado al encuestado # 6.....	43
Tabla 25. Cuestionario de actitudes realizado al encuestado # 7.....	43
Tabla 26. Cuestionario de actitudes realizado al encuestado # 8.....	44
Tabla 27. Cuestionario de actitudes realizado al encuestado # 9.....	44
Tabla 28. Cuestionario de actitudes realizado al encuestado # 10.....	45
Tabla 29. Respuesta dada por los expertos para cada indicador	46

Introducción

Con el surgimiento de la web, también conocida como *Word Wide Web* (www), se despliega una cantidad de información que crece exponencialmente a medida que se avanza en el tiempo. La web propicia que personas separadas en el tiempo y el espacio, puedan intercambiar datos, expresar sus experiencias en diferentes idiomas; de esta manera todo tipo de información puede ser compartida digitalmente con el menor esfuerzo (Lapuente, 2013).

Debido a este cúmulo de datos que se encuentra en la red, disponible a millones de usuarios, surgen los motores de búsquedas. Estos motores tienen como objetivo la localización de la información mediante la indexación de documentos y su posterior recuperación, lo cual permite que los usuarios puedan acceder a los resultados de la búsqueda de forma rápida y efectiva (Rodríguez y otros, 2006). En la actualidad existen muchos sistemas de recuperación de información a nivel mundial, los más usados son Google, Baidu, Bing, Yahoo (Netmarketshare, 2015). En ocasiones los sitios web cubanos no son indexados por los buscadores internacionales lo que trae como consecuencia que los contenidos no estén accesibles para los usuarios que realizan la búsqueda. Por otra parte, cuando son indexados, por lo general son mal posicionados en los resultados de la búsqueda (Carrillo y otros, 2012).

Cuba se encuentra en una transformación social, motivada por el desarrollo de las Tecnologías de la Informatización y las Comunicaciones (TIC), con el objetivo de mitigar las consecuencias que produce ser un país bloqueado y muchas veces relegado su potencial y conocimiento científico, para dar paso a la información y los contenidos de otros países e instituciones. Como parte de estos esfuerzos se creó la Universidad de la Ciencias Informáticas (UCI) en el año 2002 (Ramos y otros, 2011). Dentro de su infraestructura docente-productiva se encuentra el Centro Ideoinformática (CIDI) de la Facultad 1. El centro cuenta con dos líneas de desarrollo, una dedicada a los portales web y otra encargada de realizar soluciones libres para Internet como el motor de búsqueda Orión, el cual se utiliza para obtener una determinada información de interés.

La necesidad de obtener información más actualizada, relevante y confiable es cada vez mayor, y está en correspondencia con el aumento exponencial de la web y los recursos que en ella se alojan (Torres, 2003). Esta necesidad responde al objetivo de realizar análisis estadístico sobre la información recopilada en los procesos de rastreo e indexación de los motores de búsqueda para valorar los resultados y apoyar la toma de decisiones (Romero, 2002). En el análisis exploratorio de los datos se emplean técnicas que permiten tener en cuenta criterios tan importantes como las estadísticas, tendencias, perfiles de publicación y perfiles de búsqueda de las personas que utilizan la web, y así apoyar el proceso de toma de decisiones (Castañeda

y otros, 2010).

El motor de búsqueda Orión posee los componentes fundamentales que caracterizan un buscador. Los procesos de rastreo e indexación son elementales para el funcionamiento del motor de búsqueda y producen el indexado de grandes volúmenes de información que podría ser aprovechada teniendo en cuenta el costo del *hardware* y *software* puesto a disposición para el almacenamiento. Autores como Podinovsky, Allen, Citroen y Betis han validado la importancia de la información en los procesos de apoyo a la toma de decisiones, sin embargo, todo lo que actualmente Orión indexa en su base de datos sólo se utiliza para dar respuesta a las consultas que ejecutan los usuarios, por lo que en estos momentos no se pueden tomar decisiones en cuanto a los perfiles de búsquedas, de publicación y estadísticas.

Los perfiles de búsqueda cuentan con reportes sobre búsquedas indebidas e incentivan la búsqueda de contenido dirigido a la superación profesional y la investigación, a partir de las necesidades de búsquedas de los usuarios de una región específica. Los perfiles de publicación consisten en corregir o incentivar el estilo de redacción y los contenidos de un sitio web. También tienen en cuenta el aumento de publicación de temas específicos que sean poco abordados. En el caso de las estadísticas, si se conoce el comportamiento de los usuarios y publicación en sitios web en períodos de tiempo determinados, podría deducirse comportamientos en fechas específicas y proyectar acciones intencionadas.

Por lo antes analizado se considera que no se explotan todas las potencialidades que brinda el almacenamiento de la información recopilada. Dicha información se puede utilizar y procesar con diferentes objetivos como: el análisis de tendencias y agrupamiento de estadísticas, que bien definidos y en un determinado período de tiempo podría mejorar el funcionamiento del motor de búsqueda Orión.

Actualmente Orión tiene implementado servicios como: alertas de búsquedas, publicidad y gestión de usuario pero no cuenta con un análisis estadístico para el apoyo a la toma de decisiones. Esto constituye una desventaja que impide mostrar información estadística en formato de gráficos o tablas, además de perfiles de búsquedas de un usuario y de publicación de un sitio web que sirva de apoyo al proceso de toma de decisiones que pueden tener un carácter político, económico, social y organizativo.

Sobre la base de los elementos expuestos anteriormente se formula el siguiente **problema de investigación**: ¿Cómo contribuir al procesamiento estadístico de la información almacenada en el motor de búsqueda Orión para apoyar el proceso de toma de decisiones?

Para la realización de la investigación se define como **objeto de estudio**: el procesamiento estadístico para el apoyo a la toma de decisiones en los sistemas de recuperación de información y el **campo de**

acción se encuentra enmarcado en el procesamiento estadístico en el motor de búsqueda Orión para el apoyo a la toma de decisiones.

Para dar solución al problema planteado, se define como **objetivo general**: desarrollar un módulo de procesamiento estadístico para el motor de búsqueda Orión que permita el apoyo a la toma de decisiones, utilizando el procesamiento estadístico de la información indexada.

Para dar cumplimiento al objetivo general antes mencionado, se dará cumplimiento a los siguientes **objetivos específicos**:

1. Describir el estado actual de las herramientas dirigidas al apoyo de la toma de decisiones.
2. Definir las tecnologías, herramientas y metodología a utilizar en la implementación de un módulo de procesamiento estadístico para el apoyo a la toma de decisiones en el motor de búsqueda Orión.
3. Diseñar las funcionalidades del módulo de procesamiento estadístico para el apoyo a la toma de decisiones en el motor de búsqueda Orión.
4. Implementar y validar las funcionalidades del módulo de procesamiento estadístico para el apoyo a la toma de decisiones en el motor de búsqueda Orión.

Hipótesis:

Con el desarrollo de un módulo de procesamiento estadístico que procese la información indexada por el motor de búsqueda Orión se contribuirá a la mejora del proceso de apoyo a la toma de decisiones.

Se define como **variable independiente**: módulo de procesamiento estadístico de información y como **variable dependiente**: proceso de apoyo a la toma de decisiones.

Tabla 1. Operacionalización de la variable dependiente

Fuente: (Creación del propio autor)

Variable dependiente	Descripción	Dimensión	Indicadores	Unidad métrica
Proceso de apoyo a la toma de decisiones.	Comprende el momento en el que el administrador accede a la información y realiza los reportes para tomar decisiones.	Dominio .cu (intranet cubana)	Efectividad de reporte	Totalmente de acuerdo (5), de acuerdo (4), ni de acuerdo ni en desacuerdo (3), en desacuerdo (2), totalmente desacuerdo (1).
			Nivel de apoyo	Totalmente de acuerdo (5), de acuerdo (4), ni de acuerdo ni en desacuerdo (3), en desacuerdo (2), totalmente desacuerdo (1).

Tabla 2. Operacionalización de la variable independiente

Fuente: (Creación del propio autor)

Variable independiente	Descripción	Dimensión	Indicadores	Unidad métrica
Módulo de procesamiento estadístico de información.	El mismo consiste en un software que ofrece las herramientas necesarias para el procesamiento estadístico de la información.	Dominio .cu (intranet cubana)	nulo	nulo

Para dar cumplimiento a los objetivos específicos se definieron las siguientes **tareas de investigación**:

1. Realización de un estudio sobre las tendencias en el procesamiento estadístico para el apoyo a la toma de decisión de sitios web.

2. Selección de las herramientas, estándares, metodología y tecnologías que se necesitan para implementar la propuesta de solución.
3. Elaboración de los artefactos requeridos por la metodología de desarrollo seleccionada.
4. Implementación del módulo de procesamiento estadístico para el apoyo a la toma de decisiones en el buscador Orión.
5. Procesamiento y documentación de las pruebas realizadas al módulo de procesamiento estadístico para el apoyo a la toma de decisiones en el motor de búsqueda Orión.

Para el desarrollo de las tareas científicas se han combinado diferentes métodos teóricos y empíricos de la investigación en la búsqueda y procesamiento de la información. Estos son:

Métodos teóricos:

- Histórico - Lógico: se emplea con el propósito de constatar teóricamente cómo ha evolucionado en el tiempo el procesamiento estadístico para el apoyo a la toma de decisiones en sistemas de recuperación de información, así como las herramientas y tecnologías utilizadas en el desarrollo de aplicaciones de este tipo.
- Analítico-Sintético: se emplea para el análisis de los elementos esenciales referentes a la información relacionada con los Sistemas de Recuperación de Información y procesamiento estadístico para el apoyo a la toma de decisiones.
- Inductivo-Deductivo: se emplea para arribar a razonamientos que puedan ser aplicables al problema a resolver luego de adquirir una serie de elementos referentes al procesamiento estadístico para el apoyo a la toma de decisiones en buscadores.

Métodos empíricos:

- Observación: se emplea comprender cómo se desarrolla el procesamiento estadístico para el apoyo a la toma de decisiones en buscadores a través del análisis de sistemas que lo implementan.
- Modelación: se emplea a través de los diagramas del lenguaje de modelado UML, se utilizará en la representación de las características y las relaciones entre los objetos de la solución.
- Entrevista: se emplea para obtener los requisitos funcionales de la solución propuesta y para interactuar con personal familiarizado con el buscador Orión.

El presente documento está estructurado en tres capítulos:

Capítulo 1: Fundamentación teórica del módulo procesamiento estadístico para el apoyo a la toma de decisiones en el motor de búsqueda Orión.

En este capítulo se definen conceptos importantes para la comprensión de la investigación, se describen las principales herramientas informáticas empleadas. Se realiza un estudio de las metodologías de desarrollo de software y los sistemas homólogos en cuanto a su objetivo.

Capítulo 2: Diseño del módulo procesamiento estadístico para el apoyo a la toma de decisiones para el motor de búsqueda Orión.

A lo largo de este capítulo se especifican los requisitos funcionales y no funcionales necesarios para lograr que el módulo de procesamiento estadístico para el apoyo a la toma de decisiones se integre de forma satisfactoria al motor de búsqueda Orión, los elementos fundamentales del diseño y de la arquitectura a tener en cuenta para la implementación del sistema, auxiliado por el modelado de diagramas.

Capítulo 3: Implementación y prueba del módulo procesamiento estadístico para el apoyo a la toma de decisiones para el motor de búsqueda Orión.

En este capítulo se muestra el modelo de implementación como resultado del diseño anteriormente realizado. Se describen las pruebas a utilizar y los resultados obtenidos de estas, con el objetivo de comprobar el correcto funcionamiento de los sistemas de recuperación de información y el procesamiento estadístico para la toma de decisiones.

Capítulo 1: Fundamentación teórica del módulo procesamiento estadístico para el apoyo a la toma de decisiones en el motor de búsqueda Orión.

1.1 Introducción

En este capítulo se define la fundamentación teórica que rige la investigación realizada acerca de las herramientas para el procesamiento estadístico con vista al apoyo a la toma de decisiones. Se tienen en cuenta los principales conceptos relacionados con el estudio del estado del arte sobre dichas herramientas informáticas. Se presentan las tecnologías, lenguajes, herramientas y metodología que serán empleadas en la implementación de la propuesta de solución.

1.2 Marco conceptual

Para lograr una mejor comprensión de la investigación, se abordan un conjunto de conceptos necesarios que están estrechamente relacionados con el dominio del problema.

Recuperación de la información

La Recuperación de Información (RI) trabaja en la forma en que se representa, almacena, organiza y se accede a la información (Baeza y otros, 1999). Otros autores plantean que la Recuperación de Información es dar respuesta a una pregunta formulada trayendo información relevante desde un gran archivo (Pérez y otros, 2000). La RI reside en establecer qué documentos son relevantes para satisfacer las necesidades de información de un usuario determinado, expresada en un lenguaje de consulta (Tolosa y otros, 2008). De acuerdo con las definiciones anteriormente expuestas, la autora del presente trabajo considera que la recuperación de información consiste en el proceso de obtener información que se encuentra almacenada y que es requerida por un usuario a través de técnicas, métodos, y herramientas de búsqueda.

Sistema de recuperación de la información

La diversidad de información poco estructurada y dispersa disponible en la web, dificulta el proceso de encontrar información útil que responda a las necesidades de los usuarios. Los Sistemas de Recuperación de Información (SRI) constituyen el mecanismo ideal para resolver este tipo de problemas. Estos permiten localizar y procesar cualquier contenido existente en la red, tales como textos, imágenes, videos, archivos de sonido (Andreu y otros, 2013). De acuerdo con Baeza (2005) los SRI deben de alguna manera interpretar el contenido de la información dentro de una colección de documentos y establecer con ellos, un orden de acuerdo al grado de relevancia que estos posean para las consultas de los usuarios.

Motores de búsqueda

Los motores de búsqueda son definidos por varios autores entre ellos se encuentra Real Academia Española (2014) donde se plantea que, en términos informáticos, un buscador o motor de búsqueda es un programa que permite acceder a información en internet sobre un tema determinado (Real Academia Española, 2014). Baeza (2005) define que los buscadores deben cumplir dos requisitos fundamentales, un tiempo corto de respuesta y una gran colección de documentos web disponibles en su índice. La calidad de un buscador reside en lo abundante, relevante y actualizada que sea su colección (Baeza, 2005). Por último, se estudia la definición de Varela (2011) y es la que se asume en esta investigación la cual plantea que “un motor de búsqueda es un sistema informático que busca archivos almacenados en servidores. Las búsquedas se hacen con palabras clave o con árboles jerárquicos por temas; el resultado de la búsqueda es un listado de direcciones web en los que se mencionan temas relacionados con las palabras clave buscadas” (Varela, 2011).

Toma de decisiones

Es el proceso mediante el cual se realiza una elección entre diferentes alternativas o formas posibles para resolver diferentes situaciones de la vida, estas se pueden presentar en diferentes contextos: a nivel laboral, familiar, sentimental, empresarial (Rodríguez y otros, 2015). La toma de decisiones consiste, básicamente, en elegir una alternativa entre las disponibles, a los efectos de resolver un problema actual o potencial (Garza y otros, 2012). La toma de decisiones a nivel individual se caracteriza por el hecho de que una persona hace uso de su razonamiento para elegir un curso de acción alternativo ante determinada situación que se le presenta. El hecho de involucrar una elección entre alternativas implica que debe existir más de un curso de acción posible. Si solo hay un curso de acción posible, no existe decisión (Campos, 2007). De acuerdo con las definiciones anteriormente expuestas, la autora del presente trabajo considera que la toma de decisiones es la forma de realizar una elección entre varias opciones de las que disponga un individuo.

Procesamiento estadístico

El procesamiento estadístico no refiere a una técnica, a un algoritmo o a un procedimiento particular. Es una filosofía de la optimización referida a mejoras de proceso continuas, usando una colección de las herramientas o estadísticas para datos y análisis del proceso. Es un componente dominante de las iniciativas totales de la calidad (Estévez, 2009).

Las herramientas comúnmente usadas en el procesamiento estadístico son:

- Organigramas.
- Gráficas de funcionamiento.
- Gráfica y análisis de Pareto.
- Diagramas de causa-efecto.
- Histogramas de frecuencia.
- Gráficas de control.
- Estudios de la capacidad de proceso.
- Planes de muestreo de aceptación.
- Diagramas de dispersión.

Procesamiento estadístico descriptivo

El procesamiento estadístico descriptivo representa las características de una serie de datos pertenecientes a una población o a una muestra (recogida, descripción, análisis y sumatorio de datos) (Borrego, 2008). Este procesamiento consiste en recolectar los datos, organizarlos, presentarlos, analizarlos e interpretar los resultados con el objetivo que los datos se comprendan más fácilmente y que sea más sencillo referirse a ellos (Romero, 2002). La estadística descriptiva o análisis exploratorio de datos ofrece modos de presentar y evaluar las características principales de los datos a través de tablas, gráficos y medidas resúmenes (Orellana, 2001). De acuerdo con las definiciones anteriormente expuestas, la autora del presente trabajo considera que los procesamientos estadísticos descriptivos son técnicas que describen, analizan y representan a un grupo de datos utilizando métodos gráficos y numéricos que resumen y presentan la información contenida en ellos.

Técnicas gráficas

Las técnicas gráficas son la representación en el plano de la información estadística, con el fin de obtener una impresión visual global del material representado, que facilite su rápida comprensión. Los gráficos pueden servir no solo como sustituto a las tablas, sino que también constituye por sí mismo una poderosa herramienta para el análisis de los datos, siendo en ocasiones el medio más efectivo no solo para describir

y resumir la información, sino también para analizarla (Sneiderman, 2015). De acuerdo con la definición anteriormente expuesta, la autora del presente trabajo considera que las técnicas gráficas son maneras de representar los datos de forma que brinden información a primera vista a través de una representación gráfica.

Técnicas numéricas

A diferencia de los métodos no numéricos o cualitativos, las técnicas numéricas intentan dar una medida numérica a la incertidumbre (probabilidad, factor de certeza). Estas medidas son combinadas de alguna forma hasta llegar a una medida de la confiabilidad de una o más posibles hipótesis (Morales, 2009). De acuerdo con la definición anteriormente expuesta, la autora del presente trabajo considera que las técnicas numéricas son una manera de sintetizar la información, caracterizar la muestra y resumir datos cualitativos en forma numérica o cuantitativa.

1.3 Estudio de sistemas homólogos en el ámbito internacional

En la actualidad existen diferentes aplicaciones que se utilizan para realizar análisis estadísticos, luego de un estudio fueron analizados *Google Analítico*, *Twitter*, *Internet live stats* ya que fueron los más relevantes que se encontraron, en busca de elementos y funcionalidades que pudieran ayudar a resolver el problema planteado. A continuación, se detallan algunas de las características de estos sistemas.

1.3.1 Google Analítico

Google Analítico es un servicio gratuito de estadísticas de sitios web. Ofrece información agrupada según los intereses de tres tipos distintos de personas involucradas en el funcionamiento de una página: ejecutivos, técnicos de *marketing* y *webmasters*. Tiene una interfaz muy completa de informes con gráficos desarrollados en *Adobe Flash* (Google Analítico, 2012).

Google analítico cuenta con varias funcionalidades como analítica de móviles, herramientas de análisis, analítica de contenido, informes sociales y analítica de publicidad que, teniendo en cuenta todas estas funcionalidades pueden obtener las tendencias por las que se inclinan los usuarios. A continuación, se muestra una breve descripción de algunas de las funcionalidades de google analítico.

- Herramientas de análisis: tienen un sencillo funcionamiento, es fácil de medir y entender cómo interactúan los usuarios con su sitio. Además, con solo unos clics, podrá crear rápidamente informes personalizados y segmentos de visitantes, e identificar datos importantes para compartirlos con su equipo. La analítica de contenido encuentra el valor oculto y el contenido principal de su sitio con los informes de contenido en google analítico. Muestra con qué frecuencia

visitan las personas cada una de las páginas de su sitio y con cual realizan convenciones, durante cuánto tiempo y en qué período realizan conversiones.

- Informes sociales: ayudan a medir el impacto de las redes sociales en los objetivos de su empresa y en las conversiones relacionadas. Así mismo, los datos sociales y web integrados aportan una visión integral de su contenido y comunidad.
- Analítica de publicidad: comprueba todos los canales digitales, la búsqueda, las redes sociales, los canales de afiliación y los de correo electrónico.

1.3.2 Twitter

Es una aplicación web gratuita de *microblogging* que reúne las ventajas de los blogs, las redes sociales y la mensajería instantánea. Esta forma de comunicación, permite a sus usuarios estar en contacto en tiempo real con personas de su interés a través de mensajes breves de texto a los que se denominan *Updates* o *Tweets* (Que es el Twitter, 2016).

Twitter analytics muestra de forma gráfica el número de impresiones (CPM) que no es más que el número de veces que un *tuits* aparece en el *timeline* de una persona que puede ser o no su seguidor, las interacciones (CPC) que no es más que el número total de veces que un usuario ha interactuado con alguno de nuestros *tuits* y la tasa de interacción (CTR), la cual representa el número de interacciones dividido por el número total de impresiones de todos los *tweets*. También se puede obtener datos como los *retweets*, los clics en enlaces, las respuestas, los favoritos o el número de veces que los seguidores han abierto un *tweet* para ver más detalles. Otro servicio muy práctico de la herramienta es que permite exportar los reportes del período que se elijas a través de archivos CSV o en un documento *Excel* (Araujo, 2014). Además, es un sistema de monitorización que puede resultar muy práctico. Desde allí se puede ver la evolución de los contenidos; ver el impacto de cada publicación o ver en cada mes cuáles son los aspectos que más han destacado (Funciones de Twitter, 2016).

1.3.3 Internet live stats

Internet live stats no es un directorio real, sino que el creador de la web usa estadísticas conocidas y públicas para realizar una estimación de cuánto contenido se está produciendo en cada segundo. Muestra una pantalla en la que varios números suben vertiginosamente donde se exponen todo lo que está pasando en Internet en este momento (Raya, 2016).

En este sistema se muestran estadísticas de: la cantidad de personas conectadas en internet, total de sitios, cantidad de correos que se envían cada segundo, cantidad de *tweets* enviados diariamente, cantidad

de búsquedas en *Google*, cantidad de *blog* que se escriben diariamente, cantidad de videos visitados en *YouTube*, cantidad de fotos subidas a Instagram, cantidad de usuarios activos en *Google*, *Twitter* y *Facebook* (Internet live stats, 2016). De estas herramientas para la investigación son más relevantes la cantidad de personas de todo el mundo que está conectada a internet y total de sitios publicados diariamente.

1.4 Estudios de sistemas homólogos en el ámbito nacional

En la búsqueda de sistemas de procesamiento estadístico y toma de decisiones en el ámbito nacional se hace un análisis de CUBANIC. A continuación, se muestran algunas características de este sistema.

1.4.1 CUBANIC

CUBANIC es un sitio web que brinda servicio de registro de nombres bajo el dominio .cu y es el responsable de administrar el dominio de primer nivel .cu de Cuba. El Ministerio de Informática y Comunicaciones (MIC) desarrolló una propuesta de indicadores para medir las acciones de gobierno orientadas a promover la informatización de la sociedad, que incluye la cantidad de cuentas de correo electrónico, la cantidad de cuentas de correo electrónico con salida internacional, la cantidad de puestos de trabajo con acceso pleno a Internet, el número de host, el número de sitios web cubanos, el promedio de accesos a sitios cubanos al mes. A través de CUBANIC, el administrador de dominios .cu que depende del Ministerio de Ciencia y Tecnología y Medio Ambiente, encargado de realizar estadísticas de distribución de dominio (Finquelievich, 2003). CUBANIC es un sitio que permite generar estadística relativa tanto a los dominios .cu genérico de segundo nivel y tercer nivel, además entre sus estadísticas (CUBANIC; 2016).

A continuación, se muestra datos estadísticos del sitio generado por los dominios .cu, com.cu, edu.cu, GOB.cu, inf.cu, NAT.cu, net.cu, org.cu.

Dominios .cu y genéricos de segundo nivel registrados por CUBANIC								
Dominios:	.cu	com.cu	edu.cu	GOB.cu	inf.cu	NAT.cu	net.cu	org.cu
Total:	1437	240	33	39	24	3	7	10
Total General:	1793							

Généricos delegados a otras organizaciones				
Dominios:	co.cu	sld.cu	tur.cu	cult.cu
Total:	3650	344	545	383
Total General:	4922			

Figura 1. Estadística del portal web CUBANIC

Fuente: (CUBANIC, 2016)

Luego de la investigación realizada sobre los sistemas de procesamiento estadístico y apoyo a toma de decisiones en el ámbito internacional y nacional, se arriba a la conclusión que para el desarrollo del módulo de procesamiento estadístico para el apoyo a la toma de decisiones en el motor de búsqueda Orión se utilizaran algunas funcionalidades de estos sistemas:

En el módulo de procesamiento estadístico para el apoyo a la toma de decisiones se toma la idea de la manera que se realizan las estadísticas en *Internet live stats* y se aplica para conocer de los usuarios los sitios recuperados y las categorías solicitadas. *Twitter* realiza el procesamiento estadístico mostrando en las gráficas el número de impresiones, las interacciones y la tasa de interacción de todos los tweets, se toma esa idea en el desarrollo del módulo de procesamiento estadístico para el apoyo a la toma de decisiones en el motor de búsqueda Orión para mostrar en las gráficas publicación de categorías y publicación de sitios web. Google analítico analiza las tendencias por las que se inclinan los usuarios se toma esa idea en el desarrollo del procesamiento estadístico para el apoyo a la toma de decisiones en el motor de búsqueda Orión.

1.5 Selección del entorno de desarrollo para la construcción de la solución

Para el desarrollo de la propuesta de solución se hace necesario la definición de las herramientas, tecnologías y metodología que se emplearán en correspondencia con las usadas en el desarrollo del proyecto Orión.

1.5.1 Metodología de desarrollo de software

La metodología para el desarrollo de *software* es un modo sistemático de realizar, gestionar y administrar un proyecto para llevarlo a cabo con altas posibilidades de éxito. Una metodología para el desarrollo de *software* comprende los procesos a seguir sistemáticamente para idear, implementar y mantener un *software* desde que surge la necesidad del producto hasta que se cumple el objetivo por el cual fue creado (Somerville, 2005).

La metodología de desarrollo, en ingeniería de *software*, es un conjunto de herramientas, técnicas, procedimientos y soporte documental encaminados a estructurar, planificar y controlar el proceso de desarrollo de forma organizada y lógica, que tiene como objetivo apoyar a los desarrolladores en la creación de un nuevo *software* (Zambrano, 2013).

Se puede decir que la metodología de desarrollo es un proceso de construcción del *software* con el objetivo de trazar una línea secuencial de las actividades que se deben realizar, haciendo uso de métodos,

modelos, herramientas y gestión de las aptitudes del equipo de desarrollo para que todo esto aporte a la calidad final del producto.

Para el desarrollo de la solución se usará la variación de la metodología de Proceso Unificado Ágil (AUP por sus siglas en inglés), por ser la empleada en el centro CIDI en el desarrollo de proyectos.

El Proceso Unificado Ágil o *Agile Unified Process* (AUP) en inglés es una versión simplificada del Proceso Unificado Racional (RUP). Este describe de una manera simple y fácil de entender la forma de desarrollar aplicaciones de *software* de negocio usando técnicas ágiles y conceptos que aún se mantienen válidos en RUP. La UCI le ha realizado modificaciones con el fin de adaptarlo al ciclo de vida definido para la actividad productiva de dicha institución; de las 4 fases que encierra la metodología AUP se simplificaron a (Ambler, 2014):

- Inicio: en esta fase se realiza un estudio inicial de la organización cliente que permite obtener información fundamental acerca del alcance del proyecto, realizar estimaciones de tiempo, esfuerzo, costo y decidir si se ejecuta o no el proyecto.
- Ejecución: en esta fase se recogen las actividades que desarrolla AUP de elaboración, construcción y transición. Se ejecutan las actividades requeridas para desarrollar el software, incluyendo el ajuste de los planes del proyecto considerando los requisitos y la arquitectura.
- Cierre: En esta fase se analizan tanto los resultados del proyecto como su ejecución y se realizan las actividades formales de cierre del proyecto.

Ventajas de AUP:

- Simplicidad: apuntes concisos.
- Agilidad: procesos simplificados del RUP.
- Centrarse en actividades de alto valor: esenciales para el desarrollo.
- Herramientas independientes: a disposición del usuario.
- Fácil adaptación de este producto: de fácil acomodo (HTML).

1.5.2 Marco de Trabajo

Los marcos de trabajo o *framework* son de vital importancia para el desarrollo de sistemas de software orientado a objetos a gran escala. Prometen una mayor productividad y el tiempo de salida al mercado más corto a través del diseño y la reutilización del código. Un framework es una colección organizada de clases que constituyen un diseño reutilizable para un dominio específico de software. Contiene un conjunto de librerías, componentes de software y directrices arquitectónicas que ofrece al desarrollador un *kit* de

herramientas completo para construir una aplicación de principio a fin, siempre teniendo en cuenta que es necesario adaptarlo a cada aplicación en particular (GNU, 2015).

Symfony

Symfony es un marco de trabajo que posee la característica de que es fácil de instalar y puede ser usado tanto en Linux como en Windows. Es compatible con varios sistemas gestores de base de datos como son PostgreSQL, MongoDB. Este marco de trabajo tiene una baja curva de aprendizaje, es decir, que es fácil aprender a trabajar con él y permite construir aplicaciones robustas. Incluye herramientas adicionales que ayudan a probar, depurar y documentar el proyecto (Symfony, 2015). Teniendo en cuenta que la interfaz web del buscador cubano Orión está desarrollada haciendo uso del marco de trabajo Symfony, el cual aporta beneficios a los usuarios por su facilidad de adaptación a los casos más complejos (Potencier, 2011) se selecciona este marco de trabajo para el desarrollo del módulo de procesamiento estadístico para el apoyo a la toma de decisiones. La versión del marco de trabajo que será utilizará es 2.8

1.5.3 Lenguajes

Los lenguajes de programación son el idioma mediante el cual se comunican las máquinas entre ellas y con los seres humanos. A continuación, se presentan los lenguajes que serán utilizados para la realización de la solución tanto del lado del cliente como del lado del servidor, seleccionados debido a que son los que emplea el marco de trabajo Symfony y haciendo uso de las tecnologías que utiliza Orión.

Del lado del servidor

Las tecnologías en el lado del servidor se encargan básicamente de atender las solicitudes de los clientes. Es donde se realiza la mayor parte del procesamiento de la información solicitada, generando una respuesta que facilite la comprensión por el usuario.

PHP

Hypertext preprocessor (PHP por sus siglas en inglés) es un lenguaje multiplataforma, de alto nivel e interpretado, compatible con la mayoría de los gestores de base de datos que se utilizan en la actualidad, es un lenguaje de programación del lado del servidor utilizado para generar páginas dinámicas. Se distribuye bajo licencia libre, por lo que se presenta como una alternativa de fácil acceso para todos. No requiere definición de tipos de variables y cuenta con manejo de excepciones (García, 2010). La versión de PHP con que será desarrollada la aplicación será la 5.5.9.

La sintaxis de PHP es similar a la del lenguaje C lo cual permite su rápida ejecución y consumiendo poca memoria, además cualquiera con experiencia en lenguajes del estilo C podrá entender rápidamente PHP.

Está compuesto de un sistema principal, un conjunto de módulos y una variedad de extensiones de código (Mariño, 2008).

Del lado del cliente

Para el trabajo en la interfaz web se utilizarán HTML, *JavaScript* y CSS como lenguaje de programación del lado del cliente, integrados todos en el marco de trabajo *Bootstrap*.

Bootstrap

Bootstrap es un *framework* desarrollado y liberado por *Twitter* que tiene como objetivo facilitar el diseño web. Permite crear de forma sencilla webs de diseño adaptable, es decir, que se ajusten a cualquier dispositivo y tamaño de pantalla y siempre se vean igual de bien. Es de código abierto, por lo que se puede usar de forma gratuita y sin restricciones. Los diseños creados con *Bootstrap* son simples, limpios e intuitivos, esto da agilidad a la hora de cargar y de adaptarse a otros dispositivos. El marco de trabajo trae varios elementos con estilos predefinidos fáciles de configurar: botones, menús desplegables, formularios incluyendo todos sus elementos e integración con *jQuery* para ofrecer ventanas e información sobre herramientas dinámicas (Domínguez, 2016).

HTML

HTML, acrónimo de *Hypertext Markup Language* (Lenguaje de marca de hipertexto), es un lenguaje de publicación especificado como un estándar por el W3C (*World Wide Web Consortium*) que permite la creación de páginas web (World Wide Web Consortium, 2014).

HTML es un lenguaje de marcas hipertextuales, diseñado para estructurar textos para generar páginas web. Se le pueden añadir características, etiquetas y funciones adicionales para el diseño de páginas web, generando un producto rápido y sencillo. La interpretación de las etiquetas es realizada por el navegador web. El lenguaje HTML es muy sencillo y permite describir hipertexto, es decir, texto presentado de forma estructurada y agradable, con enlaces que conducen a otros documentos o fuentes de información relacionadas. HTML 5 es la última versión de HTML y se utiliza para el desarrollo de este sistema, es una versión que contiene todos los elementos de sus versiones anteriores y mejora la compatibilidad entre los navegadores, dispositivos móviles y plataformas (Guiu, 2015).

CCS

CSS es un lenguaje de hojas de estilos creado para controlar la presentación de los documentos electrónicos definidos con HTML y XHTML. Es la mejor forma de separar los contenidos y su presentación, siendo imprescindible en la creación de páginas web complejas. Este lenguaje es utilizado para definir el

aspecto de todos los contenidos, es decir, el color, tamaño y tipo de letra de los párrafos de texto, la tabulación con la que se muestran los elementos de una lista y la separación entre titulares y párrafos (Navajas, 2012). La versión de CSS que se utiliza para el desarrollo del módulo procesamiento estadístico para el apoyo a la toma de decisiones es CSS3.

JavaScript

JavaScript es un lenguaje con muchas posibilidades, utilizado para crear pequeños programas que luego son insertados en una página web y en programas más grandes, orientados a objetos mucho más complejos. Con este lenguaje se pueden crear diferentes efectos e interactuar con los usuarios. Es un lenguaje que puede ser utilizado por profesionales y para quienes se inician en el desarrollo y diseño de sitios web. No requiere de compilación ya que el lenguaje funciona del lado del cliente, los navegadores son los encargados de interpretar estos códigos. Tiene la ventaja de ser incorporado en cualquier página web, puede ser ejecutado sin la necesidad de instalar otro programa para ser visualizado. *JavaScript* es soportado por la mayoría de los navegadores como *Internet Explorer*, *Netscape*, *Opera* y *Mozilla Firefox* (Mozilla Project, 2013).

1.5.4 Servidor web

A continuación, se describe el servidor web que será utilizado para la implementación de la propuesta de solución.

Nginx

Un servidor web de código abierto de alto rendimiento. Ofrece bajo consumo de recursos del sistema, gestiona la cache y el balanceo de carga. Una de las características más importantes de *Nginx* es que se trata de un software que es asíncrono y esto trae consigo la escalabilidad provocando que se gestionen las peticiones en muy pocos hilos, reduciendo las posibilidades de sobrecarga en el servidor (Velasco, 2015). La versión de *Nginx* que se utiliza para el desarrollo del módulo procesamiento estadístico para el apoyo a la toma de decisiones es 1.9.11.

Otras características que ofrece el servidor *Nginx* son:

- Capaz de manejar más de 10.000 conexiones simultáneas con un uso bajo de memoria.
- Balanceo de carga, distribuye la carga entre los servidores que formen parte de la estructura, redirigiendo cada vez la petición hacia aquella máquina que tenga una menor carga.
- Alta tolerancia a fallos.
- Compatible con el nuevo estándar de direcciones IPv6.

Se decide utilizar *Nginx* como servidor web, ya que además de poseer las características presentadas anteriormente es el utilizado en el desarrollo de la interfaz web del buscador Orión.

1.5.5 Herramientas

A continuación, se describen las herramientas que serán utilizadas para la implementación de la propuesta de solución.

Entorno de desarrollo Integrado

Como entorno de desarrollo integrado o también conocido como IDE por sus siglas en inglés (*Integrated Development Environment*) se utilizará *PhpStorm*.

PhpStorm es un potente IDE para programadores desarrollado por JetBrains. Es uno de los entornos de programación más completos de la actualidad, permite editar código no sólo del lenguaje de programación *PHP* como lo indica su nombre. Este IDE posee una excelente integración con marcos de trabajo como *Symfony 2*, permite la ejecución de códigos de la consola en su mismo entorno y reconoce la sintaxis de varios lenguajes, entre ellos, *HTML*, *CSS*, *JavaScript*, *PHP*, *Twig*. Se hará uso de la versión 8.0 (JetBrains, 2014).

Dado que JetBrains *Company* brinda la posibilidad de adquirir una licencia de código abierto para proyectos no comerciales de código abierto la cual es válida por un año se decidió utilizar este IDE en la presente investigación.

Herramientas CASE Visual Paradigm

Las herramientas CASE (*Computer Aided Software Engineering*, Ingeniería de Software Asistida por Ordenador) se utiliza para una rápida construcción de aplicaciones. Permite construir diagramas como el diagrama de proceso del negocio y el diagrama de despliegue. Se considera muy completa y fácil de usar, con soporte multiplataforma (Díaz, 2009).

El empleo de herramientas CASE permite integrar durante todo el proceso de ciclo de vida del proyecto aspectos como:

- Análisis de datos y procesos integrados mediante un repositorio.
- Generación de interfaces entre el análisis y el diseño.
- Generación del código a partir del diseño.
- Control de mantenimiento.

Acunetix

Acunetix comprueba los sistemas en busca de vulnerabilidades. Puede utilizarse para realizar escaneos de vulnerabilidades en aplicaciones web y para introducir pruebas de acceso frente a los problemas identificados. La herramienta provee sugerencias para mitigar las vulnerabilidades identificadas y puede utilizarse para incrementar la seguridad de servidores web o de las aplicaciones que se analizan, se utiliza la versión 9.5 (Acunetix, 2016).

Lenguaje de modelado

UML acrónimo de Lenguaje Unificado de Modelado es el lenguaje estándar especificado por el *Object Management Group* (OMG) para visualizar, especificar, construir y documentar los artefactos de un sistema, incluyendo su estructura y diseño. Utiliza un conjunto de símbolos y notaciones para representar gráficamente los diversos componentes que forman parte de la arquitectura de software. Permite el modelado de procesos de negocio y el modelado de requisitos apoyándose en el análisis orientado a objetos (Object Management Group, 2015).

Gestor de base datos

Un sistema gestor de base de datos es básicamente un sistema computarizado para guardar registros; es decir, es un sistema computarizado cuya finalidad general es almacenar información y permitir a los usuarios recuperar y actualizar esa información con base en peticiones. La información en cuestión puede ser cualquier cosa que sea de importancia para el individuo u organización; en otras palabras, todo lo que sea necesario para auxiliarle en el proceso general de su administración (Álvarez, 2007).

PostgreSQL

Es un sistema de gestión de bases de datos objeto-relacional, distribuido bajo licencia BSD y con su código fuente disponible libremente. Es el sistema de gestión de bases de datos de código abierto más potente del mercado y en sus últimas versiones no tiene nada que envidiarles a otras bases de datos comerciales, utiliza un modelo cliente/servidor para garantizar la estabilidad del sistema. Un fallo en uno de los procesos no afectará el resto y el sistema continuará funcionando (Group, 2015). Se usa este gestor de base de datos porque se asimila como tecnología del proyecto Orión.

Solr

Es un servidor de índice empresarial, que actúa como una base de datos no SQL. Una plataforma de búsqueda de código que funciona como un "servidor de búsquedas". Solr es escalable, permitiendo realizar búsquedas distribuidas y replicación de índices. Utiliza la biblioteca Java de búsqueda en su base para la indexación de texto completo y de búsqueda, y tiene como REST HTTP / XML y JSON APIs que hacen que sea fácil de utilizar desde prácticamente cualquier lenguaje de programación. Potente configuración

externa de Solr permite que sea adaptado a casi cualquier tipo de aplicación Java sin codificación, simplemente hay que utilizarlo con peticiones GET para realizar las búsquedas en el índice, y POST para agregar documentos. Se utilizará en su versión 4.10.3 ya que es la que se usa en el centro para el proyecto Orión. (Apache Solr, 2015).

Librería HighCharts

HighCharts es una librería escrita en *JavaScript* que permite la creación de gráficas. La librería ofrece un método fácil e interactivo para insertar graficas en aplicación web. La librería es compatible con todos los navegadores modernos incluyendo *iPhone/iPad* e *Internet Explorer* desde su versión 6. *Highcharts* se basa únicamente en las tecnologías de navegadores nativos y no requiere plugins secundarios como *Flash* o *Java*. Esta librería necesita sólo dos archivos de JS para ser ejecutado: el núcleo *highcharts.js* y *jQuery*, *MooTools* o marco *Prototype*. (Jhenry; 2013).

1.6 Conclusiones del capítulo

Con la investigación antes expuesta se arribó a las siguientes conclusiones:

- El análisis de varios conceptos relacionados con el campo de acción permitió un mejor entendimiento de la investigación.
- Las soluciones homólogas, se identificó que son sistemas con un alto nivel de aceptación pero cada uno responde a las necesidades particulares; aun así fue posible reconocer a partir de ellos, nuevos elementos que contribuyen al perfeccionamiento de la propuesta de solución.
- Las herramientas, tecnologías y metodología permitió obtener una base tecnológica adecuada que permitirá el desarrollo de la solución.

Capítulo 2: Diseño del módulo procesamiento estadístico para el apoyo a la toma de decisiones para el motor de búsqueda Orión

2.1 Introducción

El presente capítulo abordará el diseño del módulo procesamiento estadístico para el apoyo a la toma de decisiones para el motor de búsqueda Orión. Se identifican las clases del dominio y la relación que existe entre ellas es mostrada a través del diagrama de clases del modelo del dominio, también se identifican los requisitos funcionales y no funcionales con los que debe cumplir el sistema como vía para definir las futuras funcionalidades de la aplicación. A lo largo del capítulo se muestran los principales artefactos de ingeniería de *software* correspondientes a las funcionalidades.

2.2 Modelo del dominio

Un modelo de dominio o modelo conceptual, como también se lo conoce, es una representación visual en forma de diagrama de las clases conceptuales u objetos del mundo real que son significativos en un dominio de interés; no se trata de un conjunto de diagramas que describen clases u objetos de *software* con responsabilidades (Larman, 2004).

Descripción de clases del modelo del dominio

Usuario: Persona que realiza las búsquedas.

Buscador: Constituye una herramienta de recuperación de información en la Web.

Rastreador: Mecanismo que se encarga de recopilar los documentos en la Web.

Indexador: Mecanismo que se encarga de indexar los documentos.

Interfaz web: Constituye la vista mediante la cual se le muestran los resultados al usuario.

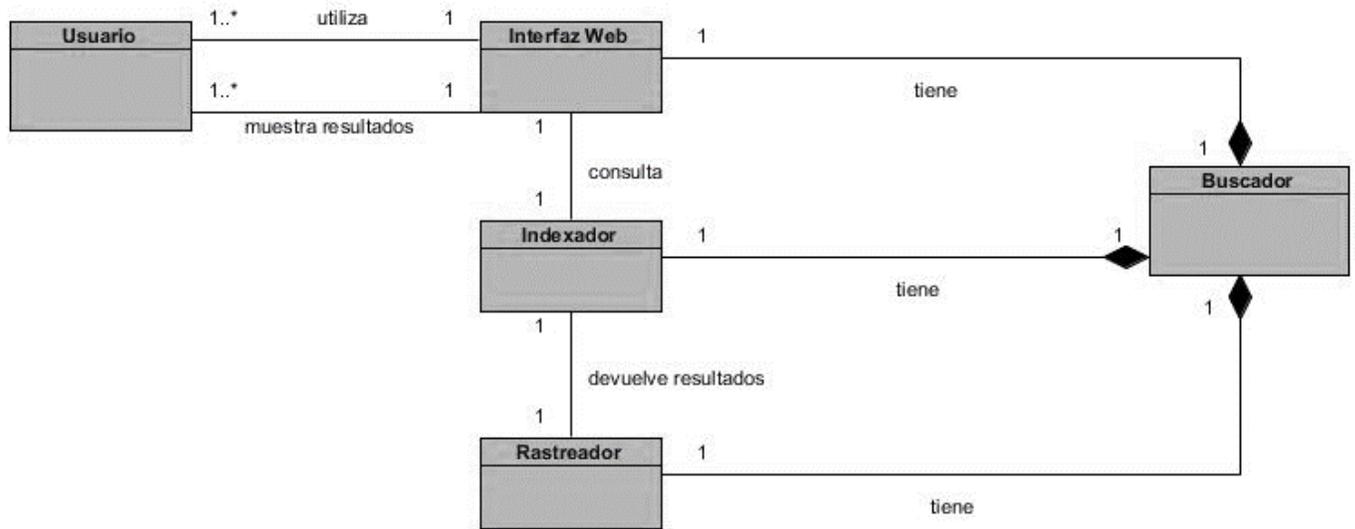


Figura 2. Modelo de dominio

Fuente: (Creación del propio autor)

2.3 Especificación de requisitos de *software*

En la ingeniería del software, los requisitos se utilizan como datos de entrada en la etapa de diseño del producto y establecen qué debe hacer el sistema, pero no cómo hacerlo. Son una condición o capacidad que un usuario necesita para poder resolver un problema o lograr un objetivo. De manera general estos requisitos son lo que el sistema debe hacer o una cualidad que el sistema debe poseer (Somerville, 2007).

2.3.1 Requisitos Funcionales

Los Requisitos Funcionales (RF) expresan la naturaleza del funcionamiento del sistema (cómo interacciona el sistema con su entorno y cuáles van a ser su estado y funcionamiento) (Somerville, 2007).

Las entrevistas realizadas arrojaron un total de 21 requisitos funcionales, que a continuación se presentan.

Tabla 3. Requisitos funcionales y no funcionales

Fuente: (Creación del propio autor)

Código	Descripción (Requisitos Funcionales)	Prioridad
RF1	Mostrar la cantidad de publicaciones de cada categoría para un sitio específico.	Alta
RF2	Mostrar la cantidad de sitios que tienen publicaciones de una categoría	Alta

	específica.	
RF3	Mostrar el comportamiento de publicación de un sitio web específico en un período de tiempo por cada categoría.	Media
RF4	Mostrar el comportamiento de una categoría específica.	Media
RF5	Mostrar el comportamiento de la categoría más publicada.	Media
RF6	Mostrar el comportamiento de la categoría menos publicada.	Media
RF7	Mostrar el comportamiento de publicación de todas las categorías en un período de tiempo específico.	Media
RF8	Mostrar la tabla de distribución de frecuencia de todas las categorías para un sitio específico.	Media
RF9	Mostrar la distribución de categorías.	Media
RF10	Mostrar los sitios menos recuperados.	Media
RF11	Mostrar los sitios más recuperados.	Media
RF12	Mostrar la categoría más solicitada por todos los usuarios.	Media
RF13	Mostrar la categoría más solicitada por un usuario.	Media
RF14	Mostrar la categoría menos solicitada por todos los usuarios.	Media
RF15	Mostrar la categoría menos solicitada por un usuario.	Media
RF16	Mostrar los días en que más buscan los usuarios.	Media
RF17	Mostrar los meses en que más buscan los usuarios.	Media
RF18	Mostrar la hora en que más buscan los usuarios.	Media
RF19	Mostrar los días en que menos buscan los usuarios.	Media
RF20	Mostrar los meses en que menos buscan los usuarios.	Media
RF21	Mostrar la hora en que menos buscan los usuarios.	Media

2.3.2 Requisitos no Funcionales

Los requisitos no funcionales (RNF) son propiedades que hacen al producto atractivo, usable, rápido y confiable. Son los encargados de darle el acabado al producto para que esté listo para usarse.

Se identificaron un total de 8 requisitos no funcionales, agrupados en 5 categorías que a continuación se presentan.

Requisitos de software

- RNF 1. Se requiere la instalación del servidor web Nginx en su versión 1.9.11 y PHP 5.5.9 o superior para poder visualizar la interfaz web.

Requisitos de hardware

- RNF 2. Para la interfaz web: 2 GB RAM, CPU de 2 núcleos y al menos 40 GB de Disco Duro.
- RNF 3. Para el servidor de base datos Solr: 2 GB RAM, CPU de 2 núcleos y al menos 70 GB de Disco Duro.
- RNF 4. Para el servidor de base datos PostgreSQL: 2 GB RAM, CPU de 2 núcleos y al menos 70 GB de Disco Duro.

Requisitos de diseño e implementación

- RNF 5. Para el desarrollo de la aplicación web se deberá utilizar *Symfony* 2.8 como marco de trabajo y PostgreSQL 9.4 como sistema gestor de base datos.

Requisitos de usabilidad

- RNF 6. Los dispositivos clientes que utilizarán la herramienta deben contar con navegadores web que soporten HTML5, CSS3 y Java Script.
- RNF 7. Se requiere el uso de herramientas y recursos de software libre, las cuales se podrán usar, modificar y distribuir libremente.

Requisitos de seguridad

- RNF 8. Los formularios deben ser protegidos por un Token de seguridad CSRF.

2.4 Historias de usuario

La Historia de Usuarios (HU) sirve para registrar los requerimientos de los clientes según el negocio y son utilizadas para poder realizar la estimación de cada una de las iteraciones durante la fase de planificación. Las HU son escritas por el equipo de trabajo en conjunto con los clientes en base a lo que se estima que es necesario para el sistema. (Balarezo, 2013).

Tabla 4. Historia de usuario #1

Fuente: (Creación del propio autor)

Historia de usuario	
Número: HU_1	Nombre: Mostrar la cantidad de publicaciones de cada categoría para un sitio específico.
Prioridad en negocio: Alta	
<p>Descripción: Permite mostrar la cantidad de publicaciones de cada categoría por un sitio específico.</p> <p>La funcionalidad comienza cuando el usuario selecciona del menú, ubicado en la parte lateral izquierda de la interfaz, la opción cantidad de publicaciones de cada categoría para un sitio específico y se le muestra en la parte derecha la gráfica.</p>	
<p>Prototipo:</p>  <p>The screenshot shows the Orion Administración web interface. The main content area displays a bar chart titled 'Publicaciones por categoría de android.uci.cu'. The chart shows the number of publications for various categories. The y-axis is labeled 'Publicaciones' and ranges from 0 to 6. The x-axis is labeled 'Categorías'. The legend includes: Sociales/Estilos de Vida, Informativos/Noticias/Medios, Arte/Entretenimiento/Cultura, Negocios/Economía, Educación, Gobierno, Tecnología, Relaciones Personales, Deporte, and Comercial/Publicitario. The chart shows 1 publication for 'Sociales/Estilos de Vida', 2 for 'Informativos/Noticias/Medios', 4 for 'Arte/Entretenimiento/Cultura', 3 for 'Negocios/Economía', and 5 for 'Educación'.</p>	

Tabla 5. Historia de usuario #2

Fuente: (Creación del propio autor)

Historia de usuario																							
Número: HU_2	Nombre: Mostrar la cantidad de sitios que tienen publicaciones de una categoría específica.																						
Prioridad en negocio: Alta																							
<p>Descripción: Permite mostrar la cantidad de sitios que tienen publicaciones de una categoría específica.</p> <p>La funcionalidad comienza cuando el usuario selecciona del menú, ubicado en la parte izquierda de la interfaz, la opción cantidad de sitios que tienen publicaciones de una categoría específica y se le muestra en la parte derecha la gráfica.</p>																							
Prototipo:																							
<table border="1"> <caption>Cantidad de sitios con publicaciones por categoría</caption> <thead> <tr> <th>Categoría</th> <th>Cantidad de Sitios</th> </tr> </thead> <tbody> <tr> <td>Sociales/Estilos de Vida</td> <td>6</td> </tr> <tr> <td>Informativos/Noticias/Medios</td> <td>7</td> </tr> <tr> <td>Arte/Entretenimiento/Cultura</td> <td>6</td> </tr> <tr> <td>Negocios/Economía</td> <td>5</td> </tr> <tr> <td>Educación</td> <td>4</td> </tr> <tr> <td>Gobierno</td> <td>5</td> </tr> <tr> <td>Tecnología</td> <td>6</td> </tr> <tr> <td>Relaciones Personales</td> <td>5</td> </tr> <tr> <td>Deporte</td> <td>5</td> </tr> <tr> <td>Comercial/Publicitario</td> <td>5</td> </tr> </tbody> </table>		Categoría	Cantidad de Sitios	Sociales/Estilos de Vida	6	Informativos/Noticias/Medios	7	Arte/Entretenimiento/Cultura	6	Negocios/Economía	5	Educación	4	Gobierno	5	Tecnología	6	Relaciones Personales	5	Deporte	5	Comercial/Publicitario	5
Categoría	Cantidad de Sitios																						
Sociales/Estilos de Vida	6																						
Informativos/Noticias/Medios	7																						
Arte/Entretenimiento/Cultura	6																						
Negocios/Economía	5																						
Educación	4																						
Gobierno	5																						
Tecnología	6																						
Relaciones Personales	5																						
Deporte	5																						
Comercial/Publicitario	5																						

Tabla 6. Historia de usuario #10

Fuente: (Creación del propio autor)

Historia de usuario	
Número: HU_10	Nombre: Mostrar los sitios más recuperados.
Prioridad en negocio: Media	

Descripción: Permite mostrar los sitios más recuperados.

La funcionalidad comienza cuando el usuario selecciona del menú, ubicado en la parte izquierda de la interfaz, la opción mostrar los sitios más recuperado y se le muestra en la parte derecha en una tabla.

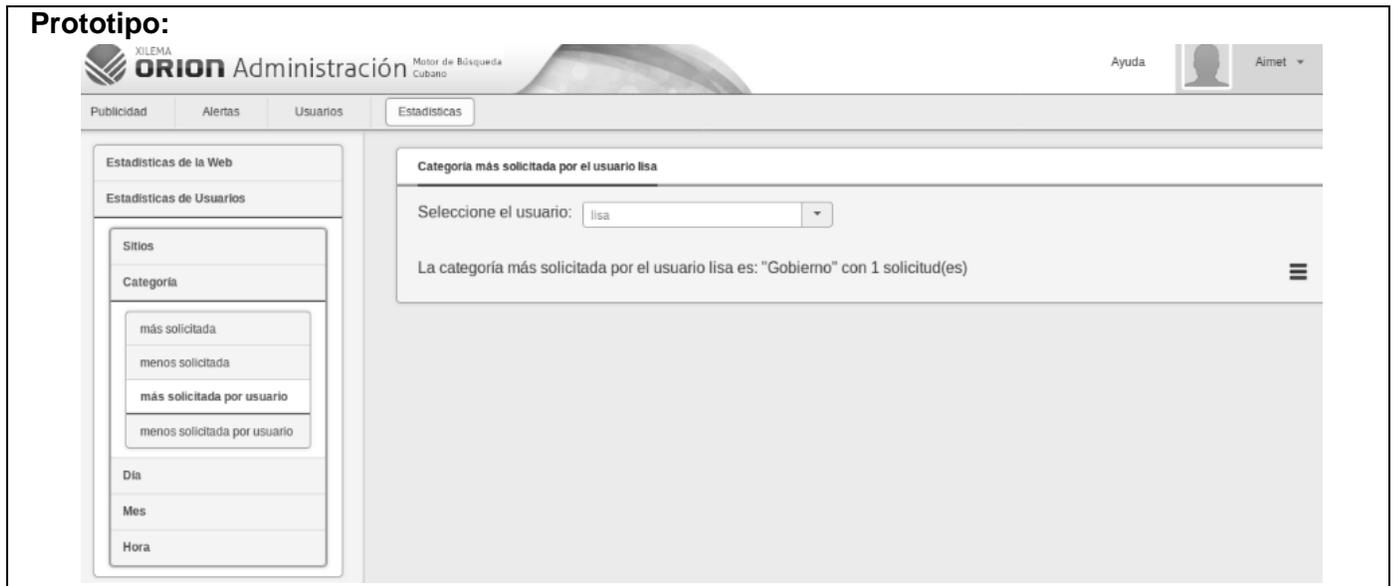
Prototipo:

The screenshot shows the Orion Administración web interface. The top navigation bar includes 'Publicidad', 'Alertas', 'Usuarios', and 'Estadísticas'. The left sidebar has a menu for 'Estadísticas de la Web' and 'Estadísticas de Usuarios', with 'Sitios' selected. Under 'Sitios', there are options for 'más recuperados' and 'menos recuperados'. The main content area shows a dropdown to 'Seleccione la cantidad a obtener: 1' and a table titled 'Sitio(s) más recuperados son:'. The table has three columns: 'No.', 'Sitio', and 'Visitas'. The data row shows '1' in the 'No.' column, 'directorio.uci.cu' in the 'Sitio' column, and '4' in the 'Visitas' column.

Tabla 7. Historia de usuario #12

Fuente: (Creación del propio autor)

Historia de usuario	
Número: HU_12	Nombre: La categoría más solicitada por un usuario específico.
Prioridad en negocio: Media	
<p>Descripción: Permite mostrar categoría más solicitada por un usuario específico.</p> <p>La funcionalidad comienza cuando el usuario selecciona del menú, ubicado en la parte izquierda de la interfaz, la opción categoría más solicitada por un usuario específico y se le muestra en la parte derecha en una tabla.</p>	



2.5 Estilo arquitectónico

Para el desarrollo del módulo procesamiento estadístico para el apoyo a la toma de decisiones se empleará como marco de trabajo *Symfony* que utiliza el patrón arquitectónico Modelo-Vista-Controlador (MVC), utilizado por la mayoría de los marcos de trabajo web. No obstante, según su creador Fabien Potencier: “*Symfony* no es un marco de trabajo MVC. Este marco de trabajo sólo proporciona herramientas para la parte del Controlador y de la Vista. La parte del Modelo es responsabilidad del usuario” (Potencier, 2011). La utilización de la arquitectura MVC en el desarrollo de la solución propuesta estableció una división lógica en tres partes principales. El modelo de datos, la presentación o interfaz de usuario (vistas) y la lógica de negocio (controladores).

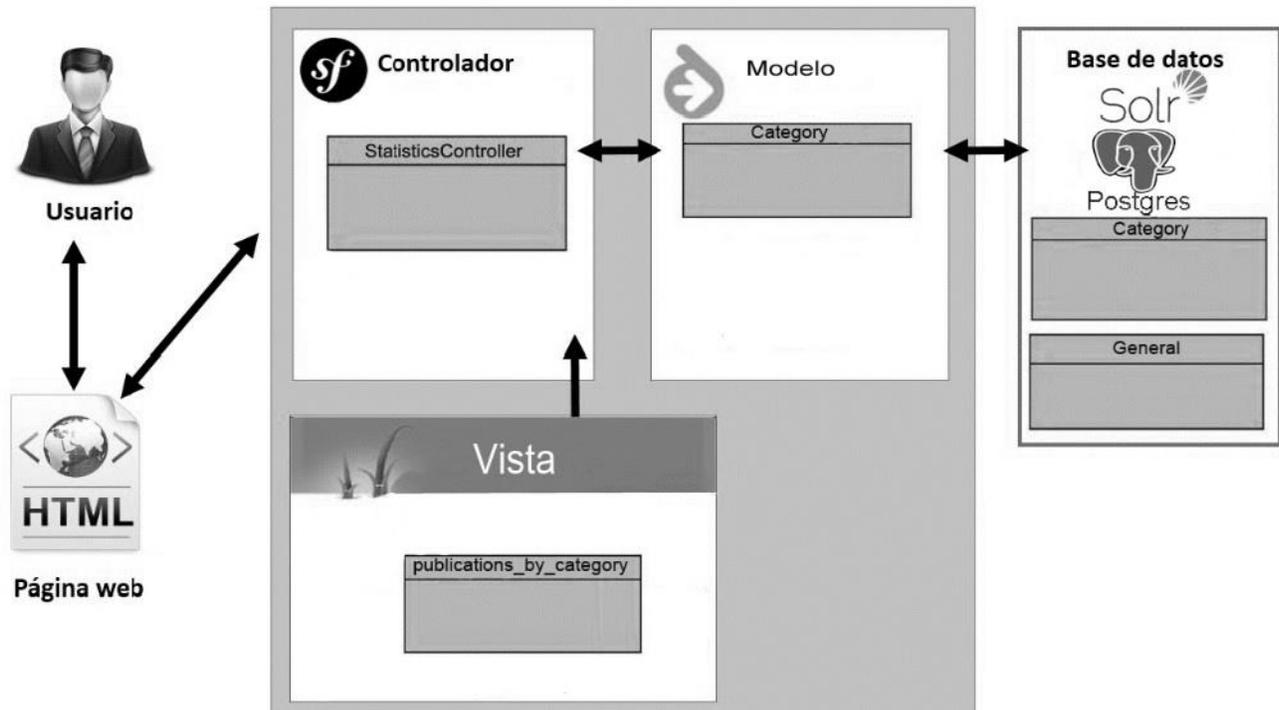


Figura 3. Arquitectura del sistema

Fuente: (Creación propia del autor)

2.6 Patrones de diseño

Un patrón de diseño es una descripción de clases y objetos comunicándose entre sí, adaptada para resolver un problema de diseño general en un contexto particular. El mismo identifica: clases, instancias, roles, colaboraciones y la distribución de responsabilidades (Wordpress, 2011).

Los Patrones Generales de Software para Asignar Responsabilidades (GRASP del inglés *General Responsibility Assignment Software Patterns*) tienen una importante utilidad en el diseño de una aplicación, al igual que los *Gang-of-Four* o Pandilla de los Cuatro (*GoF* por sus siglas en inglés). A continuación, se muestra una selección de estos patrones los cuales serán utilizados durante el diseño del módulo:

Experto

Este patrón tiene como objetivo principal asignar una responsabilidad determinada a la clase que tenga la mayor cantidad de información para hacer esta tarea. Es la razón anterior la que le da su apellido Experto

“en información”. Este patrón se evidencia en la clase *StatisticsController* es quien tiene acceso a todas las entidades necesarias y por ello a la información, por lo cual se le asigna la responsabilidad de generar todas las estadísticas que se requieren.

Creador

Se aplica para la asignación de responsabilidades a las clases relacionadas con la creación de objetos, de forma tal que una instancia de un objeto sólo pueda ser creada por el objeto que contiene la información necesaria para ello. En este caso el patrón se refleja en la clase *CategoryRepository*, encargada de crear las entidades de categoría, que será utilizado posteriormente por *StatisticsControle*.

Bajo Acoplamiento

El acoplamiento es una medida de la fuerza en que una clase está conectada a otras, que la conoce y recurre a ellas. El objetivo de este patrón consiste en mantener un bajo nivel de dependencia de otros elementos, por lo que constituye un principio que debe estar presente en todas las decisiones de diseño con lo que se reduce el impacto de los cambios.

Alta Cohesión

La cohesión es una medida de cuán relacionadas y enfocadas están las responsabilidades de una clase. Se puede afirmar que cada una de las clases del sistema tiene alta cohesión, de manera que estas poseen la característica de tener las responsabilidades estrechamente relacionadas. Esta particularidad evita en cada caso, tener que realizar un trabajo enorme al garantizar un mejor diseño en ocasiones para el resultado global.

Controlador

Para este caso se hace necesario conocer que un evento del sistema es una operación que se realiza en este, generada por un usuario externo. Un controlador, es un objeto de interfaz no destinada al usuario que se encarga de manejar un evento del sistema. Define además el método de su operación. En esta solución se encuentra ejemplo de la clase *StatisticsControle* que es la encargada de obtener todas las estadísticas para la generación de los gráficos.

Decorador

Forma parte de la familia de patrones denominados estructurales. Este tipo de patrones describen como las clases y objetos pueden ser combinados para formar grandes estructuras y proporcionar nuevas

funcionalidades. Decorador permite modificar, retirar o agregar responsabilidades a un objeto dinámicamente.

La gran ventaja es que permite extender objetos incluso en situaciones donde la extensión vía herencia no es necesaria. Adicionalmente ayuda a conservar el principio de Abierto/Cerrado, en donde se dicta que cada entidad debe estar abierta a extensión pero cerrada a modificación. Este patrón se evidencia en la entidad categoría ya que mediante anotaciones (comentarios) en los atributos de la clase se puede especificar la columna de la base de datos en la cual será guardado este atributo como se muestra en el siguiente ejemplo donde la anotación `@ORM\Column` realiza dicha función:

```
/**
 * @var string
 * @ORM\Column (name="category", type="string", length=255)
 */
private $category
```

Controlador Frontal

Es uno de los patrones de diseño de *Symfony* más conocido. Es una sección de código que atiende todas las solicitudes en la aplicación y devuelve una respuesta al navegador. Además de que este controlador se utiliza para direccionar todas las solicitudes, es el encargado de iniciar el núcleo del sistema lo que le permite decorar el mismo con características adicionales.

Uno de los principales problemas que evita es la duplicación de código, porque define un método comando en una clase abstracta, que luego se podrá redefinir en las otras clases según las particularidades que se necesiten, con el uso de decoradores. Se evidencia en los url donde todos parten del nodo `app_dev.php` ejemplo de uno de estos casos es http://localhost/orion/web/app_dev.php/publications_by_category.

2.7 Modelo de Diseño

El modelo de diseño se utiliza como medio de abstracción del modelo de implementación y el código fuente del *software*. Su objetivo fundamental es transmitir, a través de la representación mediante diagramas, una comprensión en profundidad de los aspectos relacionados con los requerimientos no funcionales y restricciones concernientes a los lenguajes de programación (Larman, 2004).

En la siguiente figura se muestra un ejemplo de Diagramas de Clases de Diseño pertenecientes al módulo procesamiento estadístico para el apoyo a la toma de decisiones para el motor de búsqueda Orión.

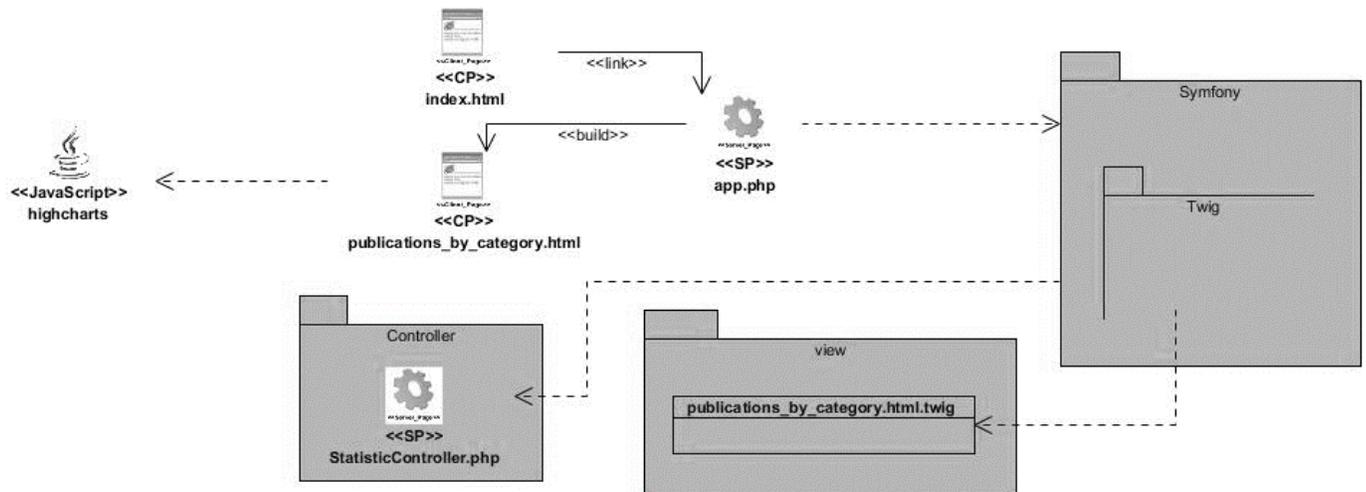


Figura 4. Diagrama de clases del diseño de publicaciones por categoría

Fuente: (Creación del propio autor)

2.8 Modelo de datos

Un modelo de datos es un sistema formal y abstracto que permite describir los datos de acuerdo con reglas y convenios predefinidos o podríamos decir que es un conjunto de conceptos que permiten describir, a distintos niveles de abstracción, la estructura de una base de datos (Redondo, 2017).

Principales características de un modelo de datos:

- Independencia lógica y física de los datos.
- Redundancia mínima.
- Respaldo y recuperación.
- Seguridad de acceso y auditoría.
- Integridad de los datos.
- Consultas complejas optimizadas.

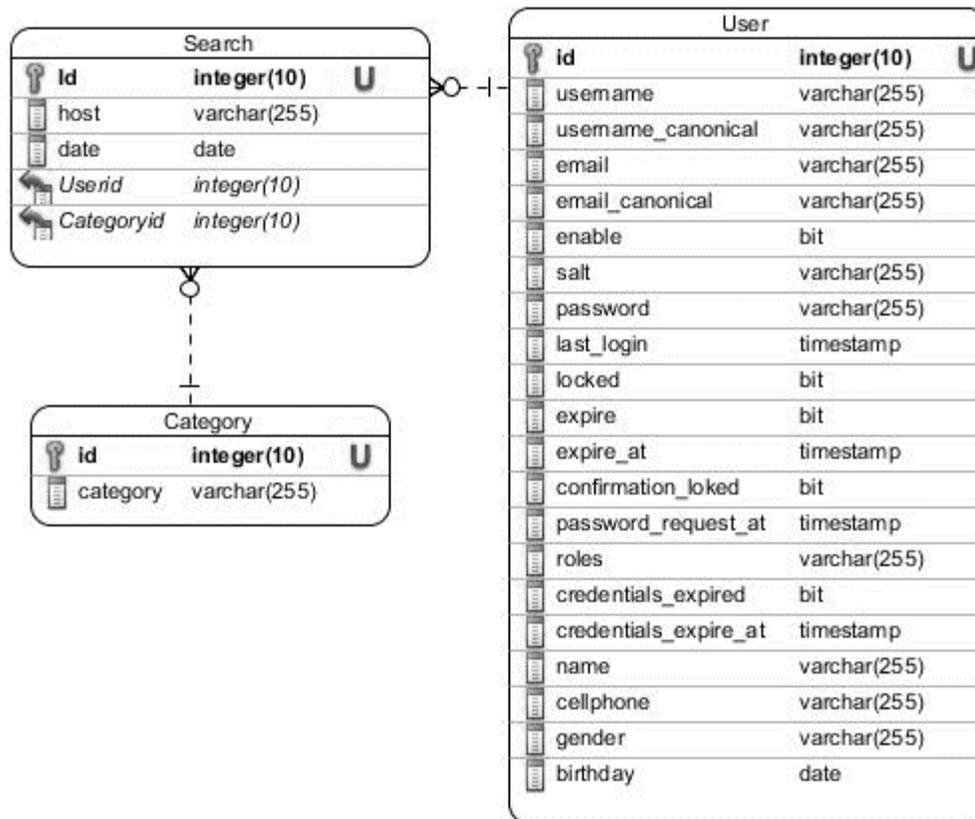


Figura 5. Modelo de datos

Fuente: (Creación del propio autor)

2.9 Modelo de despliegue

El diagrama de despliegue se utiliza para mostrar la estructura física del sistema, incluyendo las relaciones entre el *hardware* y el *software* que se despliega, estas relaciones son representadas por los protocolos de comunicación que se utilizan para acceder a cada uno (SparxSystems, 2014). En la siguiente figura puede visualizarse el diagrama de despliegue definido para la solución propuesta:

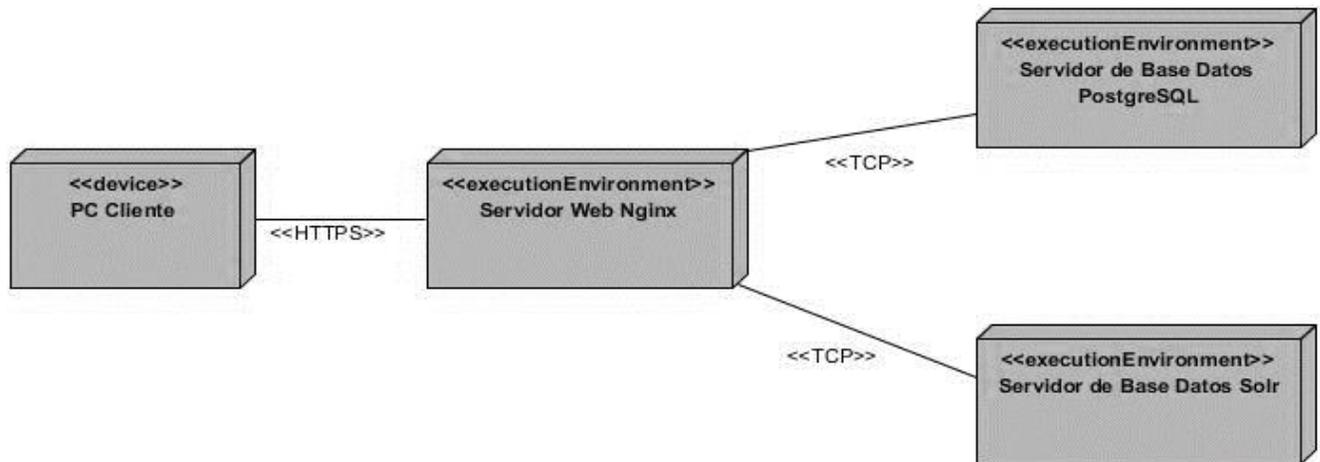


Figura 6. Diagrama de despliegue

Fuente:(Creación del propio autor)

2.10 Conclusiones del capítulo

Con la investigación antes expuesta se arribó a las siguientes conclusiones:

- Con la especificación de los requisitos funcionales y no funcionales del sistema, se logró una mejor comprensión, de los resultados que se pretenden obtener de una manera precisa y sirven de guía para la implementación del sistema.
- La representación y descripción de los artefactos generados garantizaron un mejor entendimiento de los flujos de trabajos presentes en el proceso de generar el reporte estadístico.
- La utilización de los patrones de diseño permitió identificar aspectos importantes de la estructura del diseño del sistema web propuesto, lo que garantizó una mayor organización e hizo el código más legible.
- La elaboración del diagrama de despliegue permitió identificar la disposición física de los artefactos del módulo de procesamiento estadístico para el apoyo a la toma de decisiones para el motor de búsqueda de Orión.

Capítulo 3: Implementación y validación del módulo procesamiento estadístico para el apoyo a la toma de decisiones para el motor de búsqueda Orión

3.1 Introducción

En el presente capítulo se describen los diferentes mecanismos utilizados para llevar a cabo el desarrollo y validación del sistema propuesto. Así como los estándares de codificación que debe seguir el equipo de desarrollo para implementar el software para un mejor entendimiento y organización del mismo. Luego se realiza la fase de prueba, cuyo objetivo es comprobar si el sistema cumple con los requerimientos.

3.2 Modelo de implementación

El modelo de implementación es comprendido por un conjunto de componentes y subsistemas que constituyen la composición física de la implementación del sistema. Fundamentalmente, se describe la relación que existe desde los paquetes y clases del modelo de diseño a subsistemas y componentes físicos (Hernández, 2013).

3.3 Diagrama de componentes

Un componente es una parte física de un sistema (módulo, base de datos, programa ejecutable). Se puede decir que un componente es la materialización de una o más clases (Hernández, 2013).

El diagrama de componente permite concebir el diseño según los bloques principales y ayuda al equipo de desarrollo a entender un diseño existente y crear uno nuevo. Al establecer el sistema como una colección de componentes con interfaces proporcionadas y necesarias bien definidas se garantiza la correcta separación entre los componentes. A su vez, se facilita la comprensión de los cambios al modificar los requisitos. Seguidamente se presenta el diagrama de componentes correspondiente al módulo procesamiento estadístico para el apoyo a la toma de decisiones del buscador Orión.

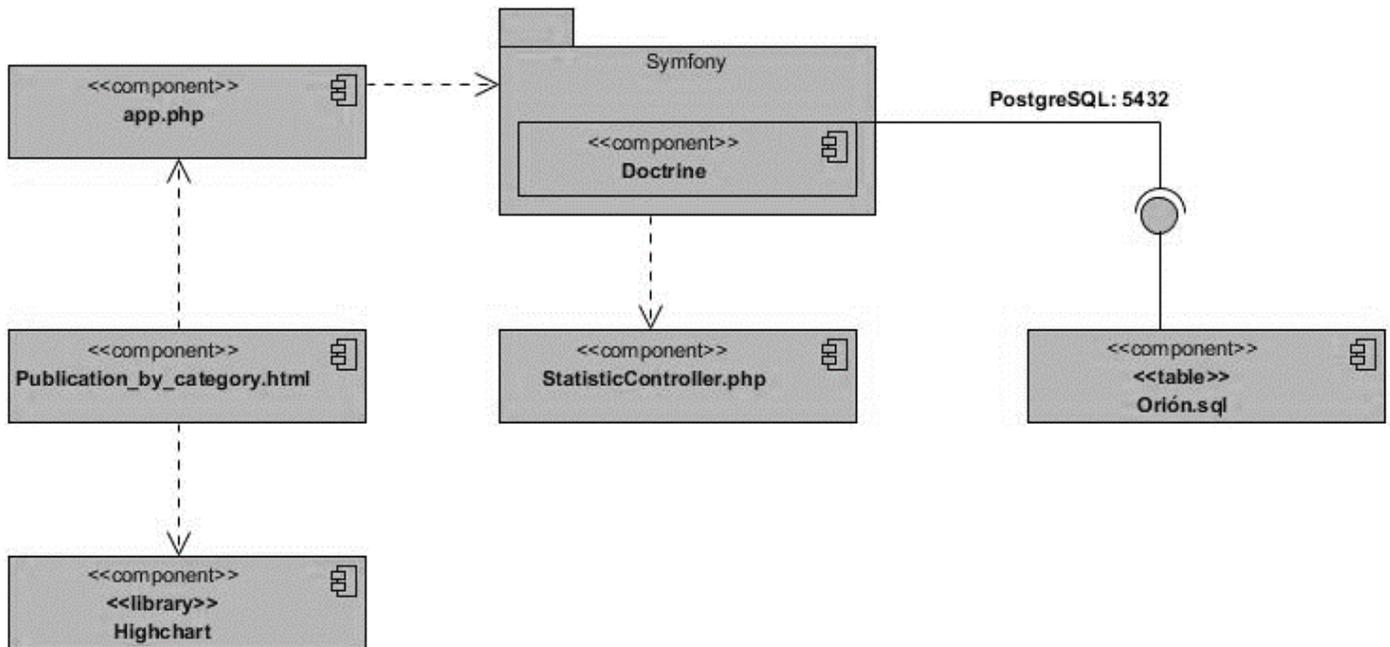


Figura 7. Diagrama de componente de publicaciones por categorías

Fuente: (Creación del propio autor)

3.4 Estándares de codificación utilizados

Un estándar de codificación completo comprende todos los aspectos de la generación de código. Un código fuente completo debe reflejar un estilo armonioso, como si un único programador hubiera escrito todo el código de una sola vez. El estándar de codificación debería establecer cómo operar con la base de código existente (Microsoft, 2017). Seguidamente se especifican los estándares de codificación a utilizar en la construcción de la solución:

- El tamaño máximo de las líneas de código debe ser de cien a ciento veinte caracteres aproximadamente, de manera tal que se garantice la completa visibilidad de las líneas de código sin necesidad de realizar desplazamiento horizontal.

```
return $this->render('StatisticBundle:web:publications_by_category.html.twig',  
    array('sites' => $sites, 'sites_availability' => $sites_availability,  
        'site_request' => $site_request));
```

Figura 8. StatisticBundle.php

Fuente: (Creación del propio autor)

- Los nombres de las clases y las funciones adoptarán la notación *UpperCamelCase* y no se utilizará el guion bajo como delimitador entre palabras.

```
public function PublicationsCategoryAction()  
{
```

Figura 9. StatisticBundle.php

Fuente: (Creación del propio autor)

- Los nombres de los atributos, variables y parámetros tendrán todas las letras en minúsculas y usarán el guion bajo como delimitador entre palabras.

```
if (is_null($site_request)) {  
    $site_request = $sites_availability[0];
```

Figura 10. StatisticBundle.php

Fuente: (Creación del propio autor)

3.5 Validación del módulo

En la ingeniería de *software* la validación es el proceso de revisión que verifica que el sistema de *software* producido cumple con las especificaciones y que logra su cometido, se trata de evaluar el sistema o parte de este durante o al final del desarrollo para determinar si satisface los requisitos iniciales.

Pruebas funcionales

Una prueba funcional es una prueba basada en la ejecución, revisión y retroalimentación de las funcionalidades previamente diseñadas para el *software*. Las pruebas funcionales se hacen mediante el

diseño de modelos de prueba que buscan evaluar cada una de las opciones con las que cuenta el paquete informático. Dicho de otro modo son pruebas específicas, concretas y exhaustivas para probar y validar que el software hace lo que debe y sobre todo, lo que se ha especificado.

El objetivo final de esta prueba es garantizar que los requerimientos hayan sido cumplidos y que el sistema goce de buena calidad. A continuación, se muestra una parte de la propuesta de diseño de casos de prueba para algunos requisitos del módulo procesamiento estadístico para el apoyo a la toma de decisiones para el motor de búsqueda Orión (Oré, 2009).

Las celdas de las tablas contienen V, I, N/A. V indica válido, I indica inválido, y N/A que no es necesario.

Tabla 8. Caso de prueba # 1

Fuente: (Creación del propio autor)

Escenario	Descripción	Sitio	Respuesta del sistema	Flujo central
EC 1.1 Mostrar la cantidad de publicaciones de cada categoría para un sitio específico.	El usuario selecciona del menú estadísticas de la web la opción publicaciones por categorías y ahí el sitio que desee mostrar la estadística.	V Directorio.uci.cu Dragones.uci.cu Android.uci.cu Humanosuci.cu	El sistema muestra de forma gráfica la opción seleccionada.	El usuario accede al menú estadísticas de la web y selecciona publicaciones por categoría. El sistema muestra gráficamente la selección del usuario. El usuario selecciona la opción exportar si es de su interés.

<p>EC 1.2 Mostrar la cantidad de publicaciones de cada categoría para un sitio específico.</p>	<p>El usuario no selecciona en el menú estadísticas de la web la opción publicaciones por categoría.</p>	<p>I</p>	<p>El sistema no le muestra de forma gráfica la opción publicaciones por categoría.</p>	<p>El usuario accede al menú estadísticas de la web y no selecciona publicaciones por categoría. El sistema no le muestra en gráfica dicho elemento que no fue seleccionado.</p>
--	--	----------	---	---

Tabla 9. Variables empleadas en el diseño del caso de prueba # 1

Fuente: (Creación del propio autor)

No	Nombre de campo	Clasificación	Valor Nulo	Descripción
1	Sitio	Estadísticas de la web	No	Permite seleccionar del menú estadísticas de la web, la opción publicaciones por categoría y ahí selecciona el sitio que desee mostrar la estadística.

Tabla 10. Caso de prueba # 2
Fuente: (Creación del propio autor)

Escenario	Descripción	Categoría	Respuesta del sistema	Flujo central
EC 1.1 Mostrar cantidad de sitios que tienen publicaciones de una categoría.	El usuario selecciona del menú estadísticas de la web la opción sitios por categorías y ahí la categoría que desee mostrar la estadística.	V Publicación por categoría.	El sistema muestra de forma gráfica la opción seleccionada.	El usuario accede al menú estadísticas de la web y selecciona sitio por categoría. El sistema muestra gráficamente la selección del usuario. El usuario selecciona la opción exportar si es de su interés.
EC 1.2 Mostrar cantidad de sitios que tienen publicaciones de una categoría específica.	El usuario no selecciona en el menú estadísticas de la web la opción sitios por categoría.	I	El sistema no le muestra de forma gráfica la opción publicaciones por categoría.	El usuario accede al menú estadísticas de la web y no selecciona sitios por categoría. El sistema no le muestra en gráfica dicho elemento que no fue seleccionado.

Tabla 11. Variables empleadas en el diseño del caso de prueba # 2

Fuente: (Creación del propio autor)

No	Nombre de campo	Clasificación	Valor Nulo	Descripción
1	Categoría	Estadísticas de la web	No	Permite seleccionar del menú estadísticas de la web, la opción sitios por categoría y ahí selecciona la categoría que quiere que se le muestre.

Tabla 12. Caso de prueba # 3

Fuente: (Creación del propio autor)

Escenario	Descripción	Sitio	Respuesta del sistema	Flujo central
EC 1.1 Mostrar el comportamiento de publicación de un sitio web específico en un período de tiempo por cada categoría.	El usuario selecciona del menú estadísticas de la web la opción comportamiento por publicaciones y ahí el sitio que desee mostrar la estadística.	Si	El sistema muestra de forma gráfica la opción seleccionada.	El usuario accede al menú estadísticas de la web y selecciona comportamiento por publicaciones. El sistema muestra gráficamente la selección del usuario. El usuario selecciona la opción exportar si es de su interés.

EC 1.2 Mostrar el comportamiento de publicación de un sitio web específico en un período de tiempo por cada categoría.	El usuario no selecciona en el menú estadísticas de la web la opción comportamiento por publicaciones.	Si	El sistema no le muestra de forma gráfica la opción publicaciones por categoría.	El usuario accede al menú estadísticas de la web y no selecciona comportamiento por publicaciones. El sistema no le muestra en gráfica dicho elemento que no fue seleccionado.
--	--	----	--	---

Tabla 13. Variables empleadas en el diseño del caso de prueba # 3

Fuente: (Creación del propio autor)

No	Nombre de campo	Clasificación	Valor Nulo	Descripción
1	Sitio	Estadísticas de la web	No	Permite seleccionar del menú estadísticas de la web, la opción comportamiento por publicaciones y ahí selecciona el sitio que desee mostrar la estadística.

Resultados de las pruebas funcionales

Para probar el correcto funcionamiento del sistema se realizaron cuatro iteraciones de pruebas. En la tabla que a continuación se presenta, se muestran los resultados obtenidos en cada iteración de pruebas al módulo de procesamiento estadístico para el apoyo a la toma de decisiones, así como la corrección de cada uno de los errores.

Tabla 14. Cantidad de no conformidades por cada iteración las pruebas

Fuente: (Creación del propio autor)

No conformidades	Primera iteración	Segunda iteración	Tercera iteración	Cuarta iteración
Detectadas	15	3	2	0

Resueltas	15	3	2	0
Pendientes	0	0	0	0

En la tabla 14 se muestra el comportamiento de las no conformidades encontradas durante el proceso de prueba donde se observa que en la primera iteración se detectaron quince (15) no conformidades, de ellas cuatro (4) errores de interfaz, tres (3) validaciones incorrectas, cuatro (4) errores ortográficos y cuatro (4) de idioma, en la segunda iteración se encontraron tres (3) no conformidades, de ellas una (1) validación incorrecta, dos (2) errores de interfaz, en la tercera iteración se encontraron dos (2) no conformidades, dos (2) errores de interfaz y en la cuarta iteración no se encontró no conformidades.

Pruebas de seguridad

Las pruebas de seguridad buscan medir la confidencialidad, integridad y disponibilidad de los datos, desde la perspectiva del aplicativo, es decir partiendo a identificar amenazas y riesgos desde el uso o interface de usuario final. Una vez ejecutadas las pruebas de seguridad es posible medir y cuantificar los riesgos a los cuales se ven expuestos los aplicativos tanto en la infraestructura interna como externa (V&V Quality, 2016).

Resultados de las pruebas de seguridad

Para garantizar la seguridad del módulo procesamiento estadístico para el apoyo a la toma de decisiones fue utilizada la herramienta Acunetix en su versión 9.5 donde se detectó 8 alertas en total, de ellas 4 de clasificación media y 4 de clasificación baja. En la primera iteración se encontraron cuatro (4) no conformidades de formularios no validados y cuatro (4) no conformidades de urls no validados.

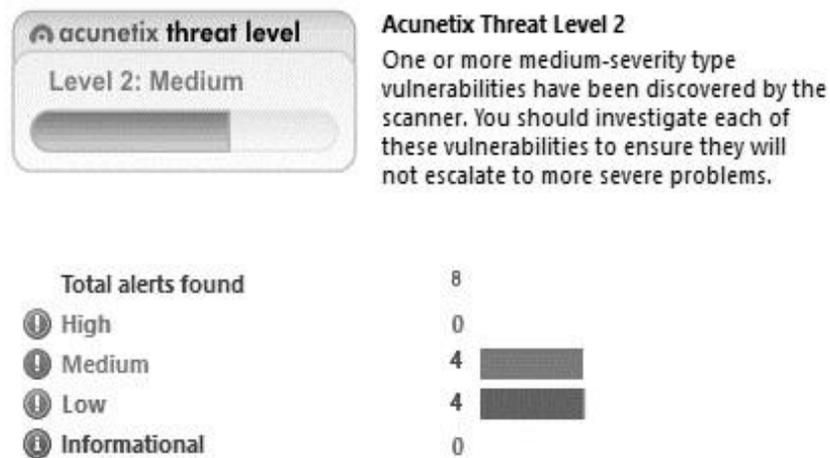


Figura 11. Prueba de seguridad iteración 1

Fuente: (Creación propia del autor)

Los ataques realizados a la aplicación se mencionan seguidamente.

- Ataques de inyección.
- Cross-Site Scripting (XSS).
- Falsificación de petición (CSRF).
- Detección de ficheros y directorios.

Después de analizar los resultados obtenido, se procedió a corregir las deficiencias encontradas y se realizó una nueva iteración en la que no se detectaron errores.

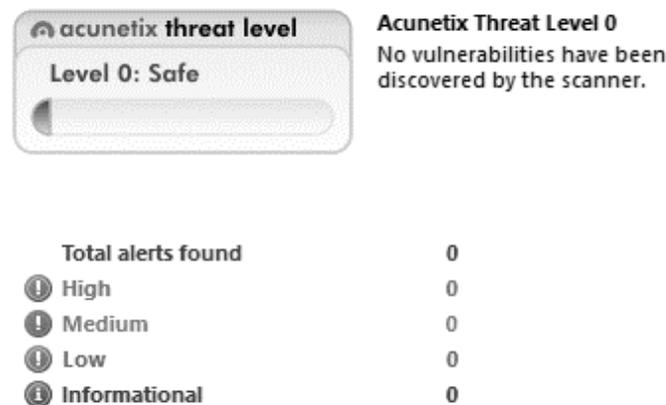


Figura 12. Prueba de seguridad iteración 2

Fuente: (Creación propia del autor)

Pruebas de usabilidad

En el contexto del desarrollo de software, la usabilidad está considerada como uno de los factores de calidad de mayor importancia para el éxito de un proyecto. De manera general, el término usabilidad es empleado para referirse a la capacidad que posee un producto de ser utilizado por los usuarios de forma fácil, eficiente y con satisfacción, en un determinado contexto de uso (Perurena y otros, 2013).

Se le aplicó la lista de chequeo UCI al módulo de procesamiento estadístico para el apoyo a la toma de decisiones del buscador Orión. Dicha lista está desarrollada por el centro de calidad UCI.

Resultados de las pruebas de usabilidad

Tabla 15. Resultados de prueba de usabilidad utilizando lista de chequeo

Fuente: (Creación del propio autor)

Categoría de Indicadores	Indicadores	Correctos
Visibilidad del sistema	17	15
Lenguaje común entre sistema y usuario	11	11
Libertad y control por parte del usuario	29	28

Consistencia y estándares	33	32
Estética y diseño minimalista	18	18
Prevención de errores	8	8
Ayuda a los usuarios a reconocer, diagnosticar y recuperarse de los errores	11	4
Ayuda y documentación	11	10
Flexibilidad y eficiencia de uso	6	6
Total	144	132

En la tabla anterior se evidencia que la implementación del módulo procesamiento estadístico para el apoyo a la toma de decisiones cumple 132 de los 144 indicadores de usabilidad lo cual representa el 91,6% aproximadamente. De las 12 no conformidades dos (2) visibilidad del sistema, una (1) libertad y control por parte del usuario, una (1) consistencia y estándares, siete (7) ayuda a los usuarios a reconocer, diagnosticar y recuperarse de los errores y una (1) ayuda y documentación.

Luego de analizar los resultados obtenidos en las pruebas de usabilidad, se identificaron que los 12 indicadores tenían posibles mejoras de acuerdo al alcance del presente trabajo. Con esto se logró un incremento en el nivel de usabilidad hasta de un 8,4% aproximadamente. En la gráfica siguiente se representa el estado del nivel de usabilidad resultante.

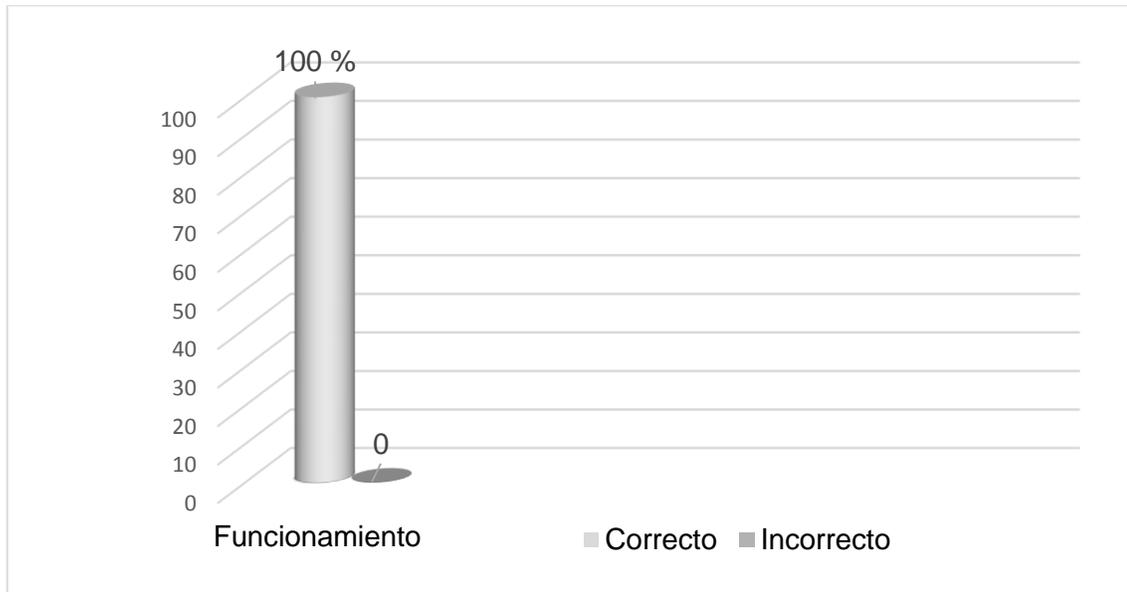


Figura 13. Nivel de usabilidad después de corregidos indicadores de usabilidad

Fuente: (Creación del propio autor)

Pruebas de integración

Este tipo de pruebas consiste en la comprobación de que los elementos del *software* que interactúan entre sí, funcionan de manera correcta. Es una forma de chequear la correcta interrelación de los distintos componentes del sistema; en el caso de la solución desarrollada es la verificación de una correcta interoperabilidad entre el módulo desarrollado y el buscador Orión.

Para ello se integró el *StatisticBundle*, que contiene el procesamiento estadístico implementado, al proyecto Orión esta operación arrojó un total de 4 errores, a continuación se muestran los resultados obtenidos en cada iteración de pruebas, así como la corrección de cada uno de los errores. En la primera iteración se encontró una (1) no conformidad que fue migrar la base de datos para PostgreSQL ya que Orión la tiene implementada en MongoDB, en la segunda iteración se encontró tres (3) no conformidades, de ellas la primera no conformidad fue adicionar el bundle *StatisticsControle* al kernel de Orión que se encuentra en `app/AppKernel.php`, la segunda no conformidad fue que no lograba la conexión con el servidor de Solr por adicionar el bundle *StatisticsControle* al fichero `routing.yml` que se encuentra en `app/config/routing.yml` y la

tercera no conformidad fue cambiarle la conexión con el servidor de Solr al fichero parameters.yml la dirección IP del servidor local del Solr que se encuentra en app/config/parameters.yml

Resultados de las pruebas de integración

Tabla 16. Cantidad de errores por cada iteración las pruebas de integración

Fuente: (Creación del propio autor)

Errores	Primera iteración	Segunda iteración	Tercera iteración
Detectados	1	3	0
Resueltos	1	3	0
Pendientes	0	0	0

3.6 Criterio de expertos

Se utiliza el método criterio de expertos por técnica de *Likert*. Esta técnica permite medir actitudes y conocer el grado de conformidad del encuestado en cuanto al cuestionario de actitudes realizado (Ospina y otros, 2003).

La entrevista se realizó a 10 expertos de ellos 6 ingenieros informáticos y 4 máster en ciencias, cada uno de estos encuestados tienen más de 3 años de vinculación en proyectos de software (Anexo 1).

Los expertos expresan sus valoraciones mediante indicadores como se presentan en la siguiente tabla:

Tabla 17. Clasificación

Fuente: (Creación del propio autor)

5	4	3	2	1
Totalmente de acuerdo	De acuerdo	Ni de acuerdo ni en desacuerdo	En desacuerdo	Totalmente desacuerdo

Los principales aspectos utilizados para aplicar la técnica de *Likert* fueron:

- Considero que se logra una interpretación de la información a través de los gráficos.
- Considero que los gráficos y la tabla de frecuencia constituyen información visual de importancia para la toma de decisiones.

- Considero que la información presentada por un período de tiempo permite identificar comportamientos.
- Considero que los datos estadísticos de los usuarios permiten entender su comportamiento.
- Considero que de forma general un usuario administrador al disponer de toda la información que brinda el módulo obtiene el apoyo necesario para la toma de decisiones.

Se realiza una tabla de valoración de expertos en cuanto al sistema para agrupar los resultados del cuestionario que se le realizó a los especialistas elegidos. Donde de estos 10 encuestados 3 de ellos estuvieron totalmente de acuerdo con las afirmaciones realizadas y los 7 restantes especialistas estuvieron de acuerdo con 1 de las preguntas y totalmente de acuerdo con las 4 restantes preguntas (Anexo 2). Con el análisis de los datos anteriormente expuesto se arriba a obtener el porcentaje de valoración de los expertos en las preguntas analizadas.

Una vez calculado los porcentaje de concordancia de las respuestas de los expertos se calcula el Índice Porcentual (IP) que integra en un solo valor la aceptación de cada planteamiento por los evaluadores mediante la siguiente fórmula:

$$IP = 5(\%) + 4(\%) + 3(\%) + 2(\%) + 1(\%) / 5$$

Tabla 18. Valoración de expertos en cuanto al sistema

Preguntas	Escala					
	IP	TD	DA	NI	ED	TA
Pregunta 1	96	80	20	0	0	0
Pregunta 2	98	90	10	0	0	0
Pregunta 3	92	60	40	0	0	0
Pregunta 4	100	100	0	0	0	0
Pregunta 5	100	100	0	0	0	0

Como se evidencia en la tabla 18 el Índice Porcentual por cada una de las preguntas es mayor que 90 lo que indica que los especialistas están satisfechos con las afirmaciones realizadas. Para un mejor entendimiento de la tabla de valoración de expertos en cuanto al sistema se muestra el siguiente gráfico.

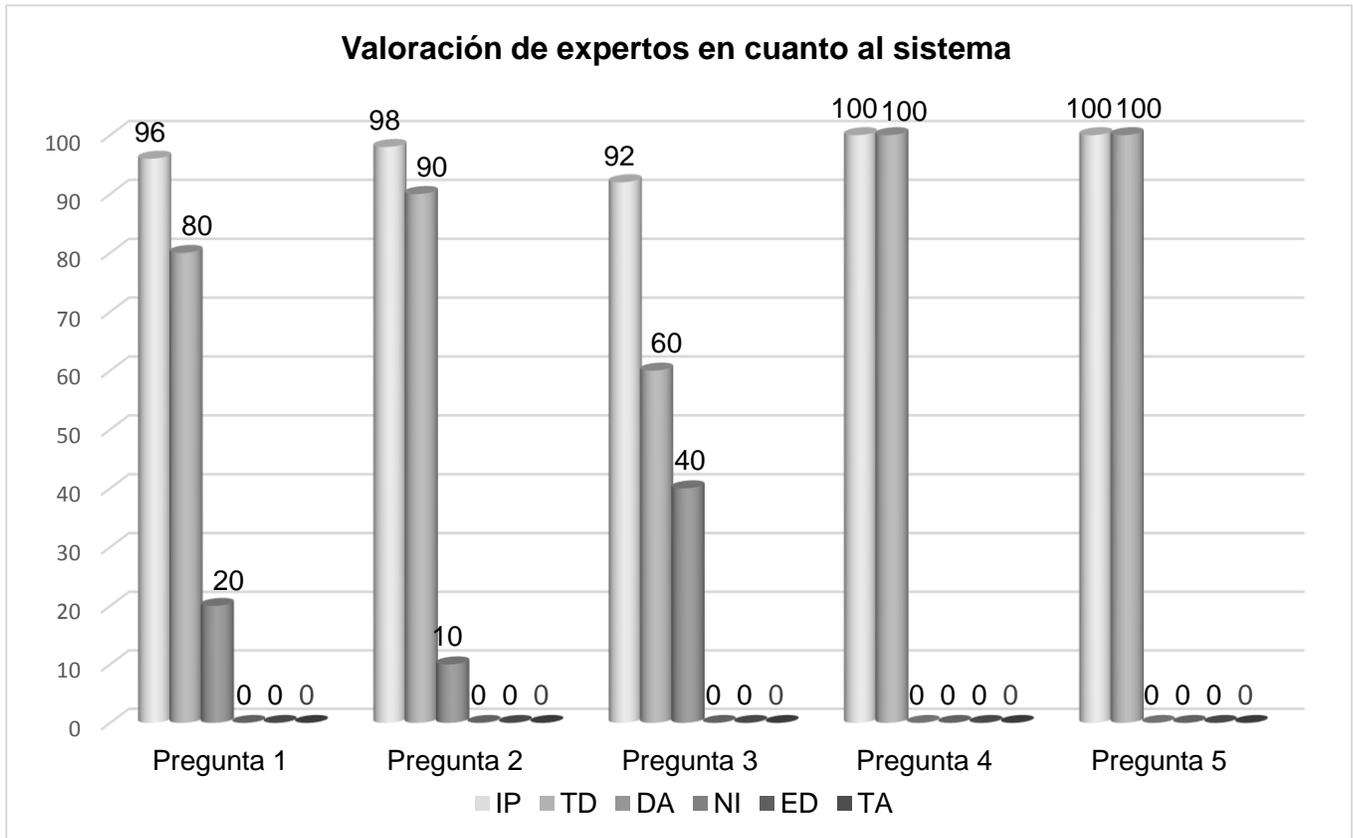


Figura 14. Distribución de frecuencia múltiple de la prueba de actitud

Fuente: (Creación del propio autor)

3.7 Conclusiones del capítulo

Con la investigación antes expuesta se concluye que:

- En el presente capítulo se especificó cómo está construido el sistema a partir del diagrama de componentes, lo cual permitió identificar con claridad la estructura y relaciones que existen entre los diferentes componentes empleados en la implementación del módulo.
- La utilización de estándares de código para la implementación de la propuesta de solución permitió adoptar una estructura homogénea que facilita la comunicación y una menor cantidad de errores; logrando un código más limpio y fácil de mantener.
- Las pruebas realizadas permitieron detectar los errores presentes, corregirlos en el menor tiempo posible y entregar al cliente una aplicación con mayor calidad, seguridad y usabilidad.

Conclusiones Generales

Una vez completada la presente investigación, se puede concluir que:

- El estudio realizado del sistema de homólogos, evidenció la necesidad de implementar un módulo de procesamiento estadístico para el apoyo a la toma de decisiones definiéndose una propuesta de solución de acuerdo a las necesidades existentes.
- La selección de herramientas, lenguajes y tecnologías permitió la implementación del módulo procesamiento estadístico para el apoyo a la toma de decisiones.
- La utilización de la estrategia de pruebas garantizó la identificación temprana de las deficiencias en el módulo que se desarrolló; corrigiéndose los mismos logrando un producto más seguro y funcional.
- El desarrollo del módulo de procesamiento estadístico para el motor de búsqueda Orión contribuyó a la mejora del proceso de apoyo a la toma de decisiones.

Recomendaciones

Una vez concluida la investigación y el desarrollo de la propuesta de solución se recomienda:

- En las vistas correspondientes a las funcionalidades de las estadísticas web dar al usuario la opción de escoger el tipo de gráfico donde se mostrarán los datos.

Referencias Bibliográficas

ACUNETIX. *Audit your website security with Acunetix Web Vulnerability Scanner.* [En línea] Acunetix, 2016. [Citado el: 28 de Noviembre de 2016]. Disponible en: <http://www.acunetix.com/>

ÁLVAREZ, S. *Sistemas gestores de bases datos.* [En línea] Introducción a este concepto y características especiales, 2007. [Citado el: 10 de Noviembre de 2016]. Disponible en: <http://www.desarrolloweb.com/articulos/sistemas-gestores-bases-datos.html>

ALVARO, J. *HighCharts: Librería para creación de gráficos.* [En línea] HighCharts: Librería para creación de gráficos, 2013. [Citado el 3 de Abril de 2017]. Disponible en: <https://enboliviacom.wordpress.com/2013/03/01/highcharts-libreria-para-creacion-de-graficos/>

AMBLER, S. *Ambyssoft.* [En línea] Ambyssoft, 2014. [Citado el: 10 de Octubre de 2016] Disponible en: <http://www.ambyssoft.com/unifiedprocess/agileUP.html>

ANDREU, V.; MANUEL, V. *Sistema de recuperación de información.* [En línea] Sistema de Recuperación de Información, 2013. [Citado el: 5 de noviembre del 2016]. Disponible en: <https://riunet.upv.es/handle/10251/19113>

BAEZA, Y.; CASTILLO, C., et al. *Crawling a Country: Better Strategies than Breadth-First for Web Page Ordering.* [En línea] Crawling a Country, 2005. [Citado el: 19 de Octubre de 2016]. Disponible en: <https://udesantiago.pure.elsevier.com/en/publications/crawling-a-country-better-strategies-than-readthfirst-for-web-p>

BAEZA, Y.; RIBEIRO, N. *Modern information retrieval.* [En línea] Modern information retrieval, 1999. [Citado el: 1 de Noviembre de 2016]. Disponible en: <http://people.ischool.berkeley.edu/~hearst/irbook/print/chap10.pdf>

BALAREZO, J. *Metodologías Ágiles.* [En línea] Metodologías Ágiles, 2013. [Citado el: 1 de Marzo de 2016]. Disponible en: <http://es.slideshare.net/EvelingGiselleCruzVs/metodologia>

BONNET, J. *Lecciones de estadística: estadística descriptiva y probabilidad,* Alicante, Club Universitario, 2004. 400 p.

- BORREGO, S.** *Estadística descriptiva e inferencial*. Innovación y Experiencias, 2008, GR 2922/2007: p. 1-12.
- CAMPOS, A.** *Pensamiento crítico: técnicas para su desarrollo*. [En línea]. Pensamiento crítico: técnicas para su desarrollo, 2007. [Citado el 14 de noviembre de 2016]. Disponible en: https://books.google.com/cu/books?hl=es&lr=&id=sMEhKEqQqR0C&oi=fnd&pg=PA9&dq=La+toma+de+decisiones+a+nivel+individual+se+caracteriza+por+el+hecho+de+que+una+persona+hace+uso+de+su+razonamiento+para+elegir+un+curso+de+acci%C3%B3n+alternativo+ante+determinada+situaci%C3%B3n+que+se+le+presenta.+El+hecho+de+involucrar+una+elecci%C3%B3n+entre+alter&ots=qM9e4_AQNK&sig=XFAx8T4w1u_wAEpogS0P-adjHU0&redir_esc=y#v=onepage&q&f=false
- CARRILLO, G.; RAMÍREZ, Y.** *Colocándonos en la web*. [En línea]. Colocándonos en la web, 2012. [Citado el: 17 de Octubre de 2016]. Disponible en <https://periodismojoosemarte.wordpress.com/2012/11/28/colocandonos-en-la-web/>
- CASTAÑEDA, M.; CABRERA, A.; et al.** *Procesamiento de datos y análisis estadísticos utilizando SPSS*. Porto Alegre, ediPUCRS, 2010. 165p.
- DÍAZ, J.** *Las metodologías ágiles como garantía de calidad del software*. Revista Española de Innovación, 2009, vol. 3: p. 40-43.
- DOMINGUEZ, M.** *¿Qué es Bootstrap y cuáles son sus ventajas?* [En línea] Bootstrap, 2016. [Citado el: 30 de Noviembre de 2016]. Disponible en: <http://puntoabierto.net/blog/que-es-bootstrap-y-cuales-son-sus-ventajas>
- ESTÉVEZ, A.** *Análisis estadísticos*. [En línea] Análisis estadísticos, 2009. [Citado el 29 de Noviembre de 2016]. Disponible en: <http://es.slideshare.net/moagmar/anlisis-estadstico>
- GARCÍA, G.** *PHP5 A través de ejemplos*. [En línea] PHP5 a través de ejemplos, 2010. [Citado el: 7 de Noviembre de 2016]. Disponible en: <http://somosprogramacion.blogspot.com/2015/02/php-5-traves-deejemplos-abraham-gutierrez-y-gines-bravo.html>
- GARZA, R.** *Concepción de un procedimiento utilizando herramientas cuantitativas para mejorar el desempeño empresarial*. Ingeniería Industrial, 2012, vol. 33: p. 239-248.
- GHEORGHIU, G.** *Agile Testing*. [En línea] Agile Testing, 2005. [Citado el 16 de Marzo de 2017]

GNU. *¿Qué es un Framework?*. [En línea] *GNUstep en español*, 2015. [Citado el: 12 de octubre de 2016] Disponible en: <https://gnustep.wordpress.com/gnustep-a-fondo/%C2%BFque-es-un-framework-%C2%BFcomo-se-utiliza>

GUIU, D. *Social Media, Empresas & TIC*. [En línea] *Social Media, Empresas & TIC*, 2015. [Citado el: 11 de noviembre de 2016]. Disponible en: <http://www.socialetic.com/que-es-html5.html>

JHENRRY, A. *HighCharts: Librería para creación de gráficos*. [En línea] *HighCharts: Librería para creación de gráficos*. [Citado el: 1 de abril de 2017]. Disponible en: <https://enboliviacom.wordpress.com/2013/03/01/highcharts-libreria-para-creacion-de-graficos/>

JETBRAINS. *PhpStorm*. [En línea] *PhpStorm*. [Citado el: 27 de Marzo de 2017]. Disponible en: <https://www.jetbrains.com/phpstorm/>

HERNÁNDEZ, L. *Modelo de implementación*. [En línea] *Modelo de implementación*, 2013. [Citado el: 15 de Marzo de 2017]. Disponible en: <http://ithleovi.blogspot.com/2013/06/unidad-5-modelo-deimplementacion-el.html>

LAPUENTE, L.; LAMARCADA, M. *Hipertexto, el nuevo concepto de documento en la cultura de la imagen*. [En línea] *Hipertexto*, 2013. [Citado el: 25 de Septiembre de 2016]. Disponible en: <http://www.hipertexto.info/documentos/web.htm>

LARMAN, C. *UML y Patrones: una introducción al análisis y diseño orientado a objetos y al proceso unificado*. [En línea] *UML y Patrones*, 2004. [Citado el: 10 de Febrero de 2015]. Disponible en: <http://www.fmonje.com/UTN/ADES%20-%202008/UML%20y%20Patrones%20%20da%20Edicion.pdf>

MARINO, C. *Programación en PHP5. Nivel básico*. [En línea] *Programación en PHP5*, 2008. [Citado el: 10 de Noviembre de 2016]. Disponible en: http://administraciondesistemas.pbworks.com/f/Manual_PHP5_Basico.pdf

MICROSOFT. *Revisiones de código y estándares de codificación*. [En línea] *Revisiones de código y estándares de codificación*, 2016. [Citado el: 15 Marzo 2017]. 2017. Disponible en: [https://msdn.microsoft.com/es-es/library/aa291591\(v=vs.71\).aspx](https://msdn.microsoft.com/es-es/library/aa291591(v=vs.71).aspx)

- MONGODB.** *Reinventando la gestión de datos.* [En línea] Reinventando la gestión de datos, 2016. [Citado el: 26 de noviembre 2016]. Disponible en: <https://www.mongodb.com/es>
- MORALES, E.** *Técnicas numéricas.* [En línea]. Técnicas numéricas, 2009. [Citado el: 10 de noviembre de 2016]. Disponible en: <https://ccc.inaoep.mx/%7Eemoraes/Cursos/Representa/node139.html>
- MOZILLA PROJECT.** *JavaScript Overview.* [En línea] Mozilla Developer Network, 2013. [Citado el: 23 de Noviembre de 2016]. Disponible en: <https://developer.mozilla.org/en-US/docs/Web/JavaScript/Overview>
- NAVAJAS, A.** *Guía completa de CSS3.* [En línea] Guía complete de CSS3, 2012. [Citado el: 22 de noviembre de 2016]. Disponible en: <https://openlibra.com/es/book/guia-completa-de-css3>
- NETBEANS.** *Bienvenido a NetBeans.* [En línea] NetBeans, 2015. [Citado el: 3 de Diciembre de 2015.] Disponible en: https://netbeans.org/index_es.html
- NETMARKETSHARE.** *Market Share Statistics for Internet Technologies.* [En línea]. Market share formobile, browsers, operating systems and search engines. 2015. [Citado el: 20 de Septiembre de 2016]. Disponible en: www.netmarketshare.com
- OBJECT MANAGEMENT GROUP.** *Introduction to OMG's Unified Modeling Language.* [En línea] Introduction to OMG's Unified Modeling Language, 2015. [Citado el: 21 de Noviembre de 2016]. Disponible en: http://www.omg.org/gettingstarted/what_is_uml.htm
- OBJECT MANAGEMENT GROUP.** *Introduction to OMG's Unified Modeling Language.* [En línea] Introduction to OMG's Unified Modeling Language, 2015. [Citado el: 23 de Noviembre de 2016]. Disponible en: http://www.omg.org/gettingstarted/what_is_uml.htm
- ORÉ, A.** *Calidad y software.* [En línea]. Calidad y software, 2009. [Citado el: 16 de Marzo de 2017]. Disponible en: http://www.calidadyssoftware.com/testing/pruebas_funcionales.php
- ORELLANA, L.** *Estadística descriptiva.* [En línea] Estadística descriptiva, 2001. [Citado el: 28 de noviembre de 2016]. Disponible en: http://www.dm.uba.ar/materias/estadistica_Q/2011/1/modulo%20descriptiva.pdf
- PÉREZ-CARBALLO, J.; STRZALKOWSKI, T.** *Natural language information retrieval: progress report. Information Processing and Management*, 2000. p. 155-178.

POTENCIER, F. *What is Symfony2?* [En línea] SensioLabsNetwork, 2011. [Citado el: 20 de Octubre de 2016]. Disponible en: <http://fabien.potencier.org/article/49/what-is-symfony2>

RAMOS, K.; et al. *Experiencias del programa de mejora de procesos en la Universidad de las Ciencias Informáticas.* Revista cubana ciencia y tecnología, 2011, 0547: p. 1-16.

REAL ACADEMIA ESPAÑOLA. *Concepto de configuración en el Diccionario de la Real Academia Española.* [En línea]. RAE, 2013. [Citado el: 5 de noviembre del 2016]. Disponible en: <http://lema.rae.es/drae/?val=informaci%C3%B3n>

REDONDO, M. *Estudio con modelos de datos para la automatización en redes eléctricas inteligentes.* [En línea]. Estudio con modelos de datos para la automatización en redes eléctricas inteligentes. [Citado el: 31 de mayo del 2017]. Disponible en: <http://hdl.handle.net/10396/14506>

RODRÍGUEZ, K.; RONDA, R. *El web como sistema de información.* Ciudad de La Habana, ACIMED, 2006. 14 p.

RODRÍGUEZ, M.; MÁRQUEZ, M. *Manejo de problemas y toma de decisiones.* [En línea]. Manejo de problemas y toma de decisiones, 2015. [Citado el 16 de noviembre de 2016]. Disponible en: https://books.google.com.cu/books?hl=es&lr=&id=yCr6CAAQBAJ&oi=fnd&pg=PT6&dq=toma+de+decision+Es+el+proceso+mediante+el+cual+se+realiza+una+elecci%C3%B3n+entre+diferentes+alternativas+o+formas+posibles+para+resolver+diferentes+situaciones+de+la+vida,+estas+se+pueden+presentar+en+diferentes+contextos:+a+nivel+laboral,+familiar,+sentimen&ots=Xw1T5fMBPF&sig=CmwNmmg2BjK79JMEzCyTkaGRfEM&redir_esc=y#v=onepage&q&f=false

ROMERO, L. *Estadística en la toma de decisiones.* República Dominicana, Universidad del Caribe, 2002. 12 p.

SNEIDERMAN, S.; GÓMEZ, J.; et al. *Expresión de patologías del desvalimiento a través de técnicas gráficas.* Técnicas gráficas, 2015, 19: p. 1-20.

SPARXSYSTEMS. *Diagrama de Despliegue UML 2.* [En línea] Sparx Systems-Tutorial UML 2-Diagrama de Despliegue, 2014. [Citado el: 15 de Febrero de 2015] Disponible en: http://www.sparxsystems.com.ar/resources/tutorial/uml2_deploymentdiagram.html

SOMERVILLE, I. *Ingeniería de Software 7ma edición*. [En línea] Ingeniería de Software, 2005 [Citado el 20 de Noviembre de 2016]. Disponible en: http://es.slideshare.net/jasc_584/ingenieriadesoftware-iansommerville7maedicion-9417118

SOMERVILLE, I. *Software Engineering* [En línea] Software Engineering, 2007. [Citado el: 5 de Marzo 2016]. Disponible en: http://eva.uci.cu/mod/resource/view.php?id=9269&subdir=/Sommerville_8va_edicion

SYMFONY. *Symfony*. [En línea] Symfony, 2015 [Citado el: 21 de Octubre de 2016]. Disponible en: <https://www.symfony-project.com>

TOLOSA, G.; BORDIGNON, F. *Introducción a la Recuperación de Información*. [En línea] Introducción a la Recuperación de Información, 2008. [Citado el: 19 de Octubre de 2016]. Disponible en: <http://admondeinformacion.blogspot.com/2011/09/introduccion-la-recuperacion-de.html>

TORRES, A. *El uso de los buscadores en Internet*. Ciudad de La Habana, ACIMED, 2003. 11p.

VARELA, O. *Búsqueda eficaz de información en la Web*. Argentina, EDULP, 2011. 126 p.

VELASCO, R. *Nginx, el servidor de páginas web comparado con Apache HTTP Server*. [En línea] Nginx, el servidor de páginas web comparado con Apache HTTP Server, 2015. [Citado el: 28 de Noviembre de 2016]. Disponible en: <http://www.redeszone.net/2015/03/29/nginx-el-servidor-de-paginas-web-comparado-con-apache-http-server>

V&V QUALITY. *V&V Quality*. [En línea] V&V Quality, 2016. [Citado el: 16 de Marzo de 2017]. Disponible en: <http://vyvquality.com/pruebas-seguridad/>

WORLD WIDE WEB CONSORTIUM. *HTML/Specifications*. [En línea] HTML/Specifications, 2014. [Citado el: 20 de Noviembre de 2016]. Disponible en: <http://www.w3.org/community/webed/wiki/HTML/Specifications#HTM>

ZAMBRANO, R. A. *Metodología de la investigación*. [En línea] Metodología de la investigación, 2013. [Citado el: 10 de Octubre de 2016]. Disponible en: <http://es.slideshare.net/Mal3J1Ta/resumen-capitulos-1234-del-libro>

ZAMORA, J; VALLEJOS, R. *Aprendiendo estadística con R.* [En línea] Aprendiendo estadística con R, 2012. [Citado el: 1 de Diciembre de 2016]. Disponible en: <http://www.cientec.or.cr/matematica/2012/ponenciasVIII/Jose-Andrey-Zamora.pdf>

Anexos

Anexo # 1: Encuestas

Tabla 19. Cuestionario de actitudes realizado al encuestado # 1

Fuente: (Creación del propio autor)

Afirmación	Alternativas de respuestas				
	1	2	3	4	5
Considero que se logra una interpretación de la información a través de los gráficos.					x
Considero que los gráficos y la tabla de frecuencia constituyen información visual de importancia para la toma de decisiones.					x
Considero que la información presentada por un período de tiempo permite identificar comportamientos.					x
Considero que los datos estadísticos de los usuarios permiten entender su comportamiento.					x
Considero que de forma general un usuario administrador al disponer de toda la información que brinda el módulo obtiene el apoyo necesario para la toma de decisiones.					x

Tabla 20. Cuestionario de actitudes realizado al encuestado # 2

Fuente: (Creación del propio autor)

Afirmación	Alternativas de respuestas				
	1	2	3	4	5
Considero que se logra una interpretación de la información a través de los gráficos.				x	
Considero que los gráficos y la tabla de frecuencia constituyen información visual de importancia para la toma de decisiones.					x
Considero que la información presentada por un período de tiempo permite identificar comportamientos.					x
Considero que los datos estadísticos de los usuarios permiten entender su comportamiento.					x
Considero que de forma general un usuario administrador al disponer de toda la información que brinda el módulo obtiene el apoyo necesario para la toma de decisiones.					x

Tabla 21. Cuestionario de actitudes realizado al encuestado # 3

Fuente: (Creación del propio autor)

Afirmación	Alternativas de respuestas				
	1	2	3	4	5
Considero que se logra una interpretación de la información a través de los gráficos.					x
Considero que los gráficos y la tabla de frecuencia constituyen información visual de importancia para la toma de decisiones.					x
Considero que la información presentada por un período de tiempo permite identificar comportamientos.					x
Considero que los datos estadísticos de los usuarios permiten entender su comportamiento.					x

Considero que de forma general un usuario administrador al disponer de toda la información que brinda el módulo obtiene el apoyo necesario para la toma de decisiones.					X
--	--	--	--	--	---

Tabla 22. Cuestionario de actitudes realizado al encuestado # 4

Fuente: (Creación del propio autor)

Afirmación	Alternativas de respuestas				
	1	2	3	4	5
Considero que se logra una interpretación de la información a través de los gráficos.					X
Considero que los gráficos y la tabla de frecuencia constituyen información visual de importancia para la toma de decisiones.					X
Considero que la información presentada por un período de tiempo permite identificar comportamientos.				X	
Considero que los datos estadísticos de los usuarios permiten entender su comportamiento.					X
Considero que de forma general un usuario administrador al disponer de toda la información que brinda el módulo obtiene el apoyo necesario para la toma de decisiones.					X

Tabla 23. Cuestionario de actitudes realizado al encuestado # 5

Fuente: (Creación del propio autor)

Afirmación	Alternativas de respuestas				
	1	2	3	4	5
Considero que se logra una interpretación de la información a través de los gráficos.					X
Considero que los gráficos y la tabla de frecuencia constituyen información visual de importancia para la toma de decisiones.					X
Considero que la información presentada por un período de tiempo permite identificar comportamientos.				X	

Considero que los datos estadísticos de los usuarios permiten entender su comportamiento.					X
Considero que de forma general un usuario administrador al disponer de toda la información que brinda el módulo obtiene el apoyo necesario para la toma de decisiones.					X

Tabla 24. Cuestionario de actitudes realizado al encuestado # 6

Fuente: (Creación del propio autor)

Afirmación	Alternativas de respuestas				
	1	2	3	4	5
Considero que se logra una interpretación de la información a través de los gráficos.					X
Considero que los gráficos y la tabla de frecuencia constituyen información visual de importancia para la toma de decisiones.					X
Considero que la información presentada por un período de tiempo permite identificar comportamientos.					X
Considero que los datos estadísticos de los usuarios permiten entender su comportamiento.					X
Considero que de forma general un usuario administrador al disponer de toda la información que brinda el módulo obtiene el apoyo necesario para la toma de decisiones.					X

Tabla 25. Cuestionario de actitudes realizado al encuestado # 7

Fuente: (Creación del propio autor)

Afirmación	Alternativas de respuestas				
	1	2	3	4	5
Considero que se logra una interpretación de la información a través de los gráficos.					X
Considero que los gráficos y la tabla de frecuencia constituyen información visual de importancia para la toma de decisiones.				X	

Considero que la información presentada por un período de tiempo permite identificar comportamientos.					x
Considero que los datos estadísticos de los usuarios permiten entender su comportamiento.					x
Considero que de forma general un usuario administrador al disponer de toda la información que brinda el módulo obtiene el apoyo necesario para la toma de decisiones.					x

Tabla 26. Cuestionario de actitudes realizado al encuestado # 8

Fuente: (Creación del propio autor)

Afirmación	Alternativas de respuestas				
	1	2	3	4	5
Considero que se logra una interpretación de la información a través de los gráficos.					x
Considero que los gráficos y la tabla de frecuencia constituyen información visual de importancia para la toma de decisiones.					x
Considero que la información presentada por un período de tiempo permite identificar comportamientos.				x	
Considero que los datos estadísticos de los usuarios permiten entender su comportamiento.					x
Considero que de forma general un usuario administrador al disponer de toda la información que brinda el módulo obtiene el apoyo necesario para la toma de decisiones.					x

Tabla 27. Cuestionario de actitudes realizado al encuestado # 9

Fuente: (Creación del propio autor)

Afirmación	Alternativas de respuestas				
	1	2	3	4	5
Considero que se logra una interpretación de la información a través de los gráficos.					x

Considero que los gráficos y la tabla de frecuencia constituyen información visual de importancia para la toma de decisiones.					x
Considero que la información presentada por un período de tiempo permite identificar comportamientos.				x	
Considero que los datos estadísticos de los usuarios permiten entender su comportamiento.					x
Considero que de forma general un usuario administrador al disponer de toda la información que brinda el módulo obtiene el apoyo necesario para la toma de decisiones.					x

Tabla 28. Cuestionario de actitudes realizado al encuestado # 10

Fuente: (Creación del propio autor)

Afirmación	Alternativas de respuestas				
	1	2	3	4	5
Considero que se logra una interpretación de la información a través de los gráficos.				x	
Considero que los gráficos y la tabla de frecuencia constituyen información visual de importancia para la toma de decisiones.					x
Considero que la información presentada por un período de tiempo permite identificar comportamientos.					x
Considero que los datos estadísticos de los usuarios permiten entender su comportamiento.					x
Considero que de forma general un usuario administrador al disponer de toda la información que brinda el módulo obtiene el apoyo necesario para la toma de decisiones.					x

Anexo # 2: Respuestas dadas por los expertos para cada indicador

Tabla 29. Respuesta dada por los expertos para cada indicador

Fuente: (Creación propia del autor)

Indicadores					
Expertos	1	2	3	4	5
1	5	5	5	5	5
2	4	5	5	5	5
3	5	5	5	5	5
4	5	5	4	5	5
5	5	5	4	5	5
6	5	5	5	5	5
7	5	4	5	5	5
8	5	5	4	5	5
9	5	5	4	5	5
10	4	5	5	5	5

Anexo # 3: Diagrama de componente

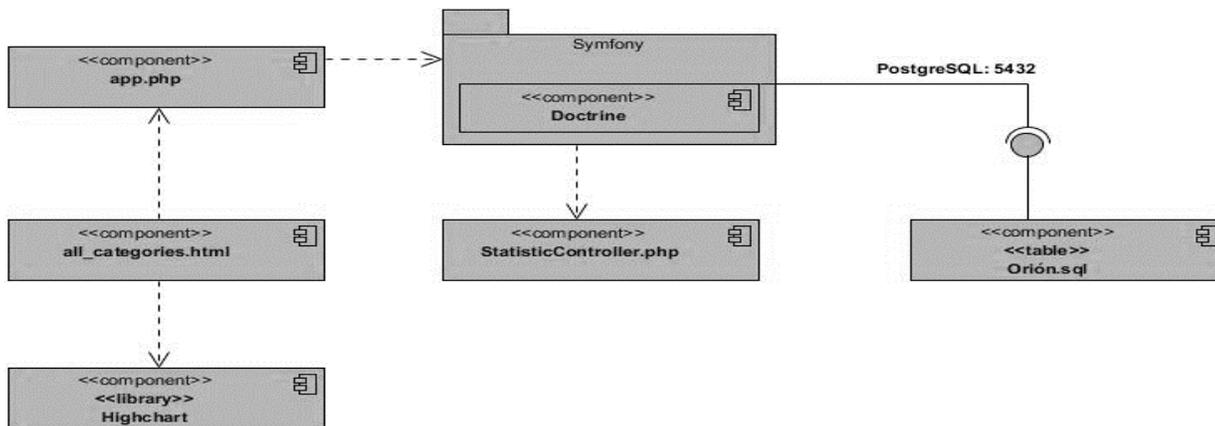


Figura 15. Comportamiento de las categorías

Fuente: (Creación del propio autor)

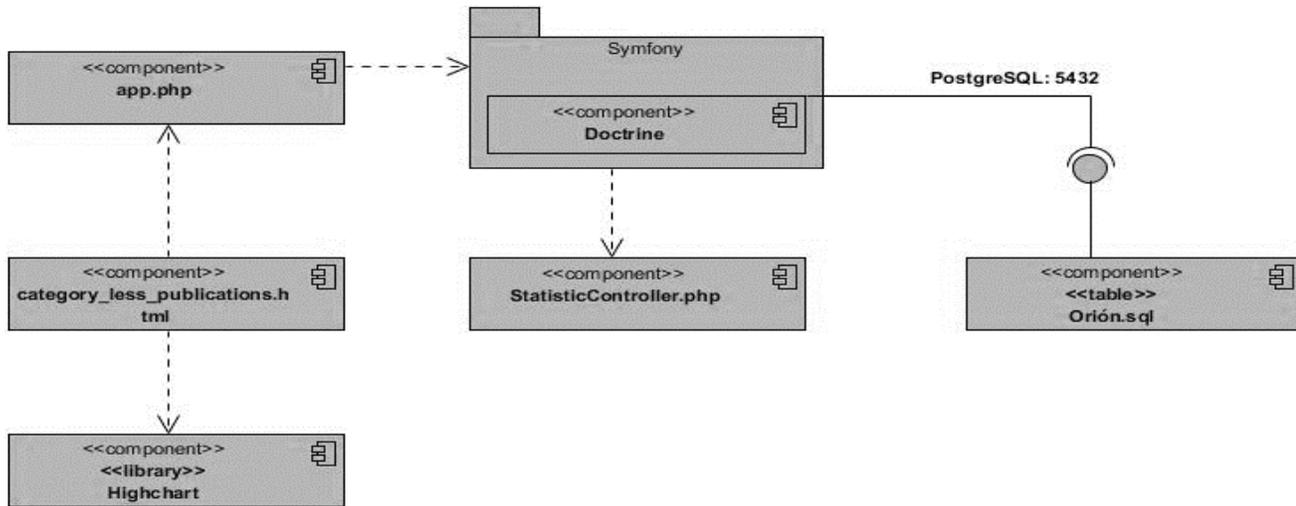


Figura 16. Categoría con menos publicaciones

Fuente: (Creación del propio autor)

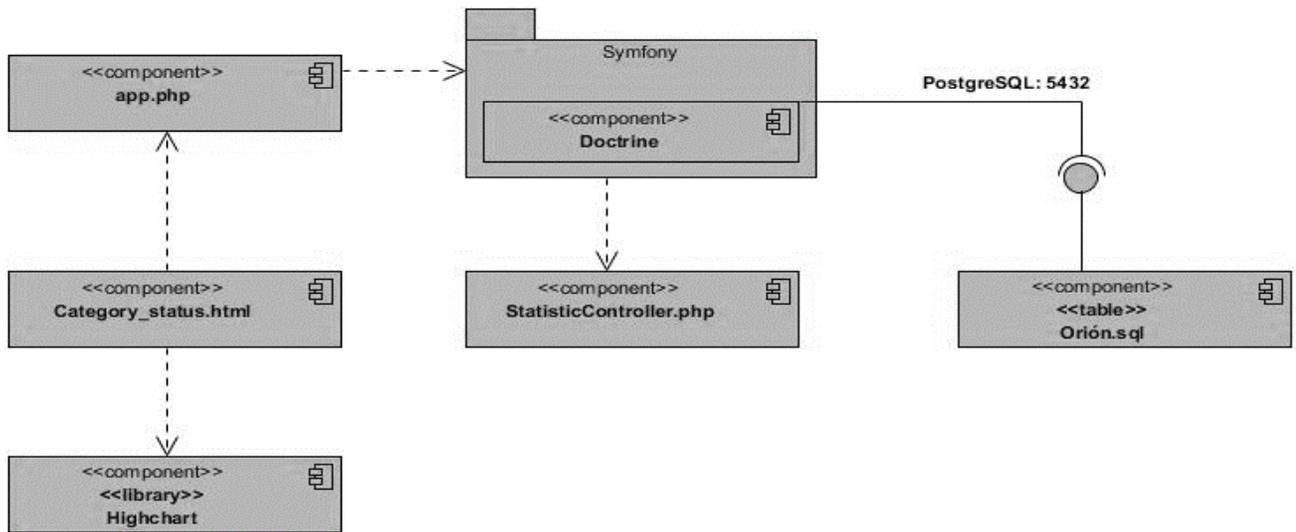


Figura 17. Publicaciones por categorías

Fuente: (Creación del propio autor)

Anexo # 4: Diagrama de diseño de clases

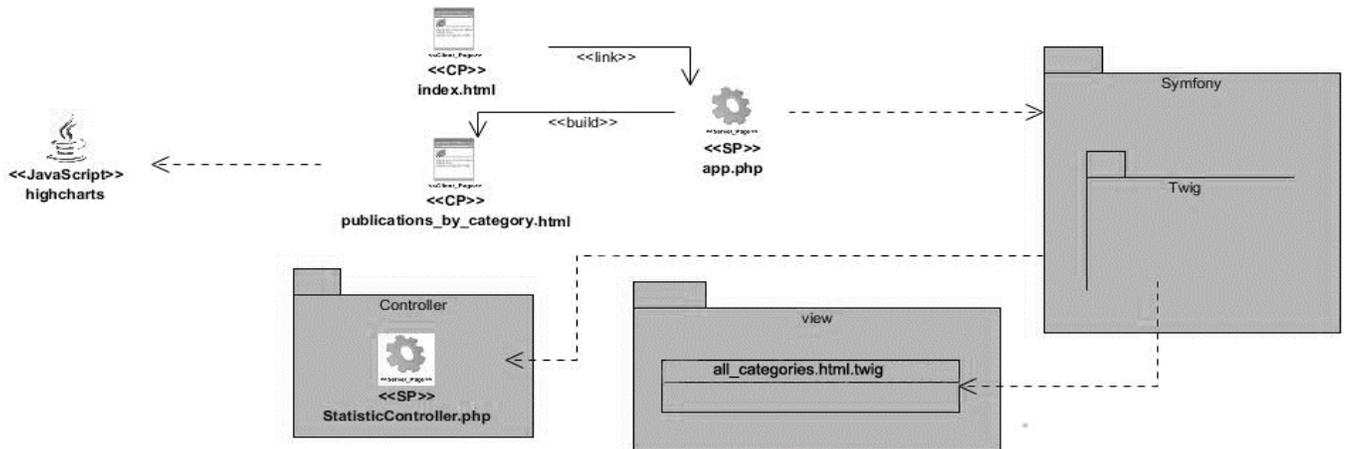


Figura 18. Comportamiento de las categorías

Fuente: (Creación del propio autor)

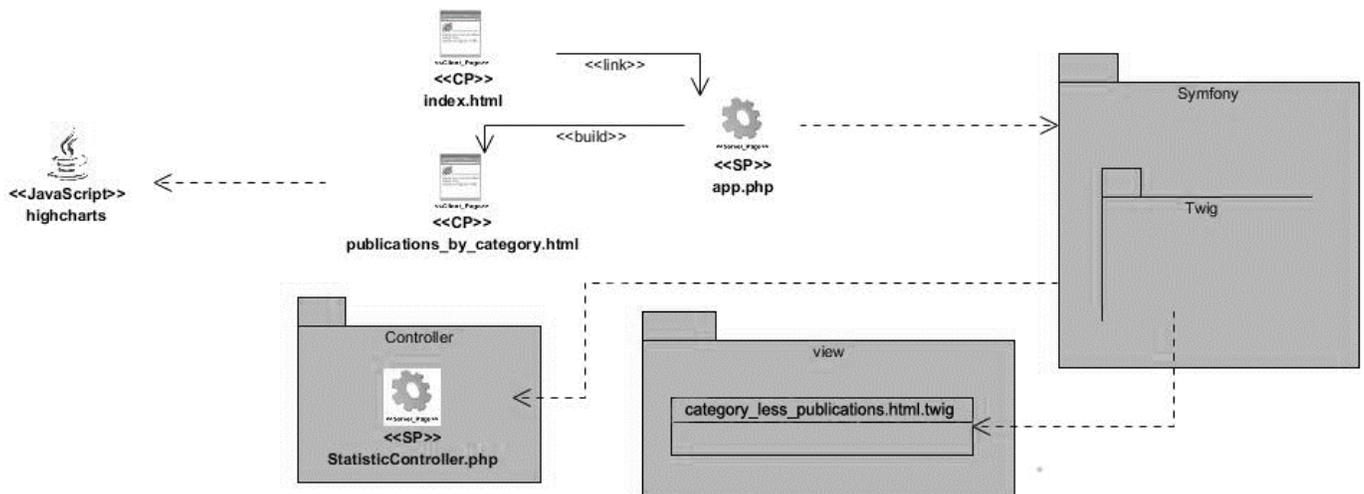


Figura 19. Categoría con menos publicaciones

Fuente: (Creación del propio autor)

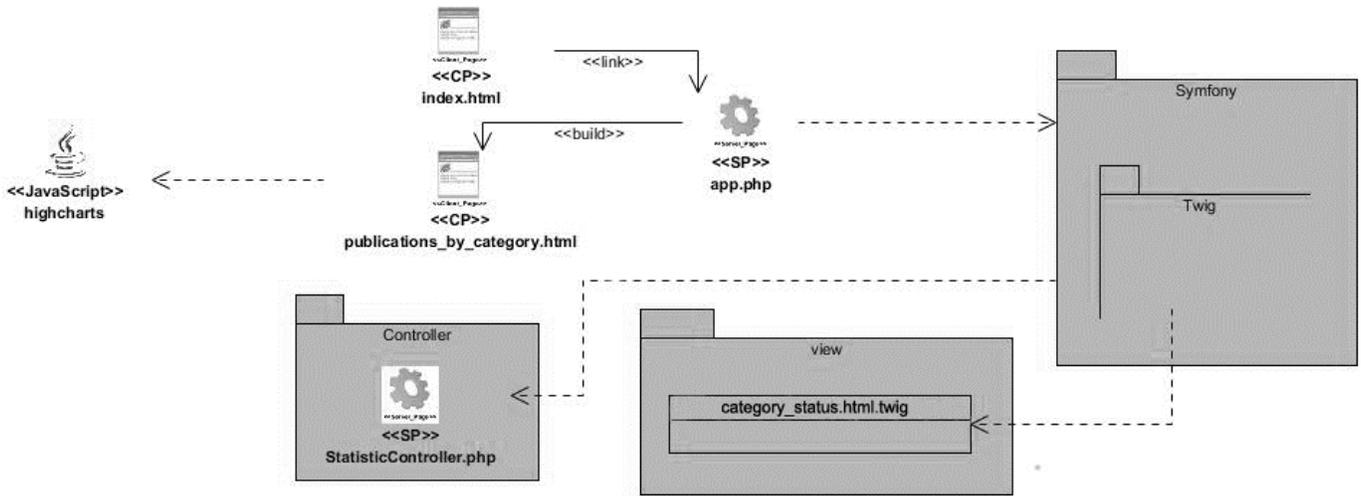


Figura 20. Publicaciones por categorías

Fuente: (Creación del propio autor)