(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 711 843**

(21) Número de solicitud: 201731290

(51) Int. CI.:

G06Q 30/00 (2012.01)

(12) SOLICITUD DE PATENTE A1

(22) Fecha de presentación:

03.11.2017

(43) Fecha de publicación de la solicitud:

07.05.2019

(71) Solicitantes:

UNIVERSIDAD REY JUAN CARLOS (100.0%)
C/ TULIPAN S/N
28933 MOSTOLES (Madrid) ES

(72) Inventor/es:

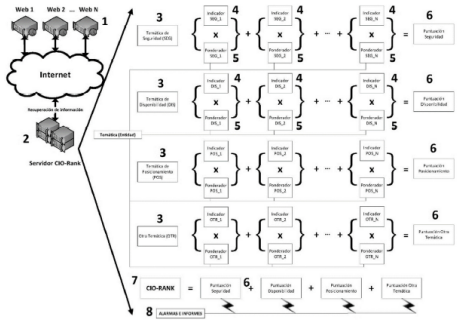
SAN MARTIN LOPEZ, Jose Javier y
SANCHEZ ZURDO, Jose Javier

(54) Título: **SISTEMA DE EVALUACION DEL COMPORTAMIENTO DE WEB**

(57) Resumen:

Método implementado por ordenador, sistema y producto de programa de ordenador para la evaluación del comportamiento de portales web, cuyo método recupera la información técnica del portal a evaluar y procesa los datos recuperados, de acuerdo a las condiciones definidas previamente en el propio método, considerando además: las temáticas de indicadores a evaluar, los indicadores dentro de cada temática, su ponderación y generando una alarma con el resultado de la evaluación.

FIG. 1



ES 2 711 843 A1

DESCRIPCIÓN

SISTEMA DE EVALUACION DEL COMPORTAMIENTO DE WEB

SECTOR DE LA TÉCNICA

La presente invención se encuadra en el área técnica de los sistemas que permiten controlar el desempeño de los sitios web. En concreto, la que atañe a la generación de alarmas de error, avisos de funcionamiento deficiente y previsión de errores futuros entre otros.

ANTECEDENTES DE LA INVENCIÓN

Actualmente la evaluación de los servicios de páginas web corporativas se orientan fundamentalmente a la consecución de un conjunto bien definido de objetivos; ya sean ventas o intangibles comerciales, tales como los retornos públicos institucionales. Sin embargo, estas estrategias de comunicación web están centradas en los intereses del emisor del mensaje, no tanto en los intereses de los receptores de las mismas. Por otro lado, solo parcialmente existen sistemas que permiten controlar y evaluar los sitios web.

Actualmente una empresa o entidad pública orienta sus estrategias de medición y monitorización de sus sistemas desde una perspectiva técnica. En algunos casos también se miden desde una perspectiva de negocio, disponiendo de indicadores que facilitan el seguimiento de los procesos clave y servicios prestados. A nivel de Administración Pública, el Observatorio de Administración Electrónica (OBSAE), publica periódicamente indicadores técnicos y de negocio como, por ejemplo, el número de transacciones con @firma, número de facturas electrónicas presentadas en FACe, número de asientos registrales (SIR), autenticaciones via Cl@ve, gasto en TIC, etc. Sin embargo no hay un traslado directo de métricas que estén orientadas al entendimiento de cómo de eficiente, buena y segura es un sistema para un cliente/ciudadano, usuario de dicho portal de información.

Aunque existen avances en esta línea por la futura transposición de la Directiva Europea 2016/2012 del Parlamento Europeo y del Consejo que obliga a mejorar la accesibilidad de los sitios web y aplicaciones móviles en el sector público, el margen temporal para su implementación es el 21 de junio de 2021 (en varias fases y cumplimientos distintos dependiendo si son portales nuevos, antiguos, aplicaciones de móviles, etc.). Esta Directiva mejorará la accesibilidad de todos los ciudadanos, con especial énfasis en aquellos colectivos con discapacidad permanente o temporal de tipo visual, auditiva, motora, sensorial, etc.

De cara a poder adaptarse a estos cambios, los gestores de los actuales de los sitios web, necesitan conocer no sólo datos concretos de funcionamiento o de evaluación por parte de los usuarios, sino también datos verificables de rendimiento, problemas del sistema, cumplimiento de normativas, etc. Existe, por tanto, una necesidad no cubierta de un sistema que permita interpretar estos datos y a su vez inferir de ellos posibles problemas futuros aunque sea únicamente a nivel de predicción.

La presente invención propone un procedimiento que evalúa la calidad de las páginas web corporativas utilizando métricas e indicadores cuantitativos, capaz de comparar instituciones similares en base a tres criterios fundamentales: calidad de los estándares de publicación, seguridad de la plataforma utilizada para su distribución y la disponibilidad de la misma.

Por otro lado, al gestor del sitio web o a la entidad reguladora superior, el procedimiento y sistema descrito en la presente patente le proporciona un conjunto de utilidades capaces de:

- Generar alarmas de mal funcionamiento o detectar problemas basados en datos reales.
- Predecir un posible mal funcionamiento (en base a combinación de varios meta-indicadores), problemas organizativos, problemas graves de disponibilidad e incluso legales.
- Realizar la combinación de las diferentes ponderaciones de los meta-indicadores de acuerdo a la naturaleza del sitio web, por ejemplo un organismo público o bien una empresa privada. Esto hace que el tratamiento de los casos anteriores sea diferenciado, y lo que es una alarma grave en un caso no lo es en otro tipo de entidad.

La presente invención ofrece una solución al problema técnico del control de sitios web, mediante la generación de las correspondientes alarmas y señales que detectan los problemas que se están produciendo e incluso los posibles problemas que pueden suceder en base al conocimiento del valor de los indicadores propuestos.

Los solicitantes de la presente invención desconocen la existencia de antecedentes que resuelvan de forma satisfactoria la problemática expuesta.

EXPLICACIÓN DE LA INVENCIÓN

El sistema descrito en la presente invención es un método implementado por ordenador y un sistema para la evaluación del comportamiento de portales web, que realiza a modo de resumen las siguientes tareas:

- recuperar información técnica del portal web a evaluar;
- procesar los datos recuperados de acuerdo a las condiciones definidas previamente en el propio método;
- define las temáticas de los indicadores a evaluar;
- dentro de cada temática se definen y puntúan los indicadores a evaluar;
- pondera los indicadores seleccionados dentro de cada temática;
- pondera cada temática;
- genera informes de resultados para cada temática
- genera alarmas con el resultado de la evaluación del portal, incluida una valoración global: valor CIO-Rank

A la hora de controlar un sitio web, en la presente invención se ha discriminado tres tipos de indicadores distintos, en base a tres temáticas:

- Seguridad del sistema a analizar.
- Disponibilidad del sistema.
- Posicionamiento de la web a nivel global.

La descripción de cada uno de estos indicadores es la siguiente:

<u>1.-Seguridad del sistema.</u>

S1. Check certificado HTTPS. Es un indicador que evalúa el número de días restantes para que el certificado HTTPS del sitio caduque. La valoración de este indicador y en su

caso penalización no es proporcionalmente lineal al número de días, sino que penaliza exponencialmente la cercanía a la fecha de caducidad.

S2. HTTPS-SHA1-test. Valor lógico que indica si está habilitada la función resumen SHA1 en el certificado HTTPS. SHA1 es un algoritmo que permite realizar resúmenes criptográficos (hash) utilizados en aplicaciones de cifrado y firma. Sin embargo su uso está totalmente desaconsejado desde el mismo momento que se demuestra que es viable comprometer su seguridad, por lo que utilizarlo puede ser un riesgo.

S3. OpenPort-03-Result-Numero Puertos Abiertos. Número de puertos de comunicación abiertos en el mismo servidor web, y por tanto, puertas abiertas hacia un número mayor de posibles vulnerabilidades del sistema. Cuantos menos puertos mejor. Se evalúa con el 100% de los puntos de esta métrica si tiene 2 puertos abiertos, siendo la caída proporcional hasta 5 puertos. Más de 5 puertos abiertos se evalúan con un 0%.

S4. SearchEngine-03-Result-Google-Documents and Settings. Documentos publicados con información institucional privada. Número de documentos indexados en Google que contienen información de usuario o rutas relativas a ordenadores internos. Lo ideal es que no haya ningún documento indexado, asignándose un 100% en el cumplimiento. En caso contrario el valor será 0%

S5. SSLlabs-03-Result-CalificacionCertificado. Calidad del certificado SSL / TLS para el servicio https, en escala de mejor a peor calidad (A-F). Esta calificación se basa en la guía para clasificar los certificados de servidores de Qualys SSLLab.

S6. Tiene Robots.txt. Valor lógico que indica si existe el fichero robots.txt. Este fichero indica a los buscadores qué indexar y qué no, dentro de un sitio web. Es deseable que no haya nada que prohíba a los buscadores que se indexe pues supone una pista a aquellos que buscan explotar informaciones ocultas. Si el valor de este test es 0 significa que obtendrá el 100% en el cumplimiento de esta métrica, en caso contrario recibirá un 0%.

S7. TipoServidorWeb. Informa del tipo de servidor web utilizado, y si es accesible la información, la versión correspondiente. En principio no hay preferencias ni calificación de mejor o peor. Es utilizado como un indicador estadístico para saber cuántos servidores hay con software comercial u Open Source.

S8. WOT_Reputation-03-ChildSafety. Indicador de seguridad infantil. Valor expresado en tanto por ciento, que representa la calidad en los contenidos en relación a la seguridad aportada a los niños, siguiendo la clasificación que ofrece la Web Of Trust (WOT).

S9. WOT_Reputation-03-Trustworthiness. Indicador de confianza-reputación. Porcentaje de calidad en cómo de seguros y confiables son los contenidos de un sitio según WOT.

2.-Indicadores de Disponibilidad

A1. Download speed for scenario "WebScenario-Principal". Velocidad media de descarga de la página web, expresada en KB/s.

A2. NTP-HTTP. Número de segundos de diferencia entre la hora oficial española y el servidor web. Lo óptimo es que estén perfectamente sincronizados. NTP-Network Time Protocol es un protocolo que permite sincronizar la hora de diferentes sistemas. En el caso de España, esta hora viene indicada por el Real Observatorio de la Armada.

A3. NumberCloneWebsByDNS. Número de servidores que gestionan dicha página web.

A4. Check DNS Expiration. Número de días para que el dominio DNS del sitio web caduque. Esta métrica se ponderará de la misma manera que el indicador S1, en el que se penaliza de manera exponencial.

A5. Response time for step "DNS-Host" of scenario "WebScenario-Principal". Tiempo de respuesta del servidor. Número de milisegundos que tarda en cargarse la página web principal, siendo deseable por supuesto el menor tiempo posible.

A6. SLA Web. Porcentaje de disponibilidad web. El 100% es el valor óptimo, aunque un umbral hasta 99.05% es un resultado razonable, dependiendo de la criticidad de los sistemas y servicios a prestar.

A7. NumberBlackListCheck. Indicador de lista negra. Número de entradas en las que aparece el dominio o la dirección IP pública de la entidad dentro de los principales servidores que indexan dominios y direcciones IP que generan spam o están catalogados como peligrosos para la seguridad.

3.-Indicadores del Posicionamiento

SEO1. PageSpeed-Desktop. Puntuación que permite comparar cómo de optimizada está la página web para un navegador de escritorio. Este indicador está basado en los resultados que se obtienen a través de la plataforma PageSpeed Insights de Google.

SEO2. PageSpeed-Mobile. Este indicador es igual que el anterior pero analizando la página web como si estuviera ejecutándose sobre un dispositivo móvil o tablet.

SEO3. Rank2Traffic Alexa. Puntuación que permite comparar la importancia de la web a nivel mundial. Alexa es una empresa que realiza análisis del tráfico en la red, estableciendo rankings entre otros, del número visitas a sitios web, posicionamiento y estrategias SEO.

SEO4. NumberCookies. Número de cookies que el servidor envía para almacenar información de sesión del usuario.

SEO5. SearchEngine-03-Result-Bing. Resultados Bing. Número de resultados con el nombre de la entidad en una consulta mediante el buscador Bing de Microsoft.

SEO6. SearchEngine-03-Result-Google. Resultados Google. Número de resultados con el nombre de la entidad en una consulta mediante el buscador de Google.

SEO7. NumberLinks. Número de enlaces a otras páginas desde la página principal de la web a analizar. Es conveniente tener enlaces a otras páginas, pero no un número excesivo, estableciéndose que por encima de 50 enlaces debería penalizarse en la página principal.

La ponderación de los indicadores, dentro de cada temática, y de las temáticas, se realiza en función de la entidad o grupo de entidades a evaluar. Un grupo de entidades es un conjunto de páginas web que comparten un conjunto de características, objetivos o finalidades comunes, lo que les hace objeto de ser comparadas entre sí, por ejemplo, páginas web de ayuntamientos constituiría un grupo de entidades. Estos grupos se pueden definir en base a dos situaciones concretas:

- Que se posea ya conocimiento explícito de las entidades y las relaciones entre ellas, es decir, que se puedan identificar entidades de manera evidente proporcionando p.ej. un servicio similar o equivalente, lo que les hace ser competidores y tiene sentido realizar una comparativa directa entre ellas.

- Que no se posea de conocimiento explícito a priori de un conjunto de entidades entre sí, de tal manera que se tengan que utilizar técnicas de clasificación, prediciendo en base a los valores de los parámetros la clasificación más acorde para cada una de las entidades. En este caso serían de aplicación sistemas inteligentes de clasificación, regresiones, utilización de redes neuronales, segmentación en subclases, Support Vector Machines (SVN), Voronoi, etc.

Establecidos los grupos de entidades, en base a características comunes definidas a priori o por extracción de conocimiento vía sistemas inteligentes de clasificación, se puede definir la ponderación de cada indicador para cada temática. Los indicadores dentro de una temática pueden ser ponderados de dos maneras diferentes:

- Inicialmente se puede realizar una ponderación manual, como p.ej. de manera proporcional (1/X siendo X el número de indicadores de la temática). Esta ponderación puede ser ajustada por el conocimiento de expertos en base a la asignación de más importancia de ciertos indicadores para un grupo concreto de páginas webs, pudiendo justificarse que fueran asimétricas por situaciones especiales de cumplimiento legal, impacto e importancia en el grupo de entidades a analizar.
- A partir de métodos estadísticos concretos o un conjunto de ellos, utilizando un conjunto amplio de muestras para realizar una exploración analítica de los datos. En base a ello se pueden identificar cuánto de importante es cada indicador y se puede aplicar un factor de corrección acorde a las necesidades del analista de datos. A modo de ejemplo, aunque no extensivo, se puede aplicar un análisis PCA (Análisis de Componentes Principales) de tal manera que se puede reducir la dimensionalidad de los indicadores y extraer las correlaciones estadísticas para aplicar automáticamente la ponderación según el Grupo de entidades y para cada una de las temáticas anteriormente mencionadas.

El Procedimiento para la reducción de la dimensionalidad y correlación de indicadores puede ser utilizado para seleccionar la ponderación de cada indicador dentro de una temática concreta y grupo de entidades seleccionadas. Para ello se seguirían los siguientes pasos:

- Se capturaría durante el tiempo suficiente los valores para cada indicador y para cada entidad.
- Si se quiere calcular una ponderación específica para un grupo de entidades, solo se seleccionarían dichos valores para esas entidades, descartando el resto de entidades. Si se quiere calcular una ponderación global (que incluyan todas las entidades con independencia de los grupos), no se descartaría ningún dato almacenado en el sistema.
- Se aplica el método matemático específico, por ejemplo PCA, basándose en los cálculos de la matriz de correlación o en cálculos de la matriz de covarianzas.
- Con los cálculos anteriores se puede determinar los valores propios y los vectores propios, utilizándose el par de ellos para determinar la importancia de cada parámetro en dicha temática (pesos)
- Se aplican dichos pesos para cada parámetro, realizando una ponderación para que el resultado de la suma de todos ellos esté ajustado entre 0 y 100%.
- Se establece un periodo temporal de validez de estos resultados. Pasado este periodo, se recalcularán los mismos si así lo estima conveniente el analista de datos.

Conocidos los indicadores y sus temáticas asociadas, grupo al que se aplican y las ponderaciones a aplicar, se definen ponderaciones a aplicar en los resultados para cada temática. Al igual que en los puntos anteriores, la manera de ponderar una temática para su evaluación global CIO-Rank podrá darse de dos maneras fundamentales:

- Realización de una ponderación manual para cada una de las Temáticas indicadas, inicialmente de manera proporcional (1/3 del valor para cada Temática). Sin embargo el analista de datos puede ajustar esta ponderación en base al conocimiento de las entidades analizadas, grupos concretos de páginas webs o como análisis previo con el cambio sustancial de las ponderaciones de indicadores para extraer conocimiento oculto utilizando técnicas “What If” o en castellano “Qué pasaría sí”.
- También se pueden utilizar métodos estadísticos concretos o un conjunto de ellos, utilizando un conjunto amplio de muestras para realizar una exploración analítica de los datos. En base a ello se pueden identificar cuánto de importante es cada temática y se puede aplicar un factor de corrección acorde a las necesidades del analista de datos. A modo de ejemplo, aunque no extensivo, se puede aplicar un análisis PCA tal como se ha explicado anteriormente.

El método para la ponderación puede ser similar al explicado en el punto anterior referido a los indicadores dentro de una temática.

De acuerdo a los valores que se obtengan en los indicadores, se genera una alarma cuando se superen los umbrales de seguridad definidos en cada indicador. La captura de la información en la que se basa cada uno de los indicadores referenciados y futuros, se almacenan de manera continua en el sistema. Es el analista de datos el que establece inicialmente un conjunto de umbrales superiores e inferiores que al sobrepasarse desencadenan acciones asociadas. Es viable utilizar más de dos umbrales, de tal manera que se utilicen técnicas de definición de funciones definidas a trozos que delimiten claramente donde se establecen cada uno de los umbrales y la criticidad de los mismos.

Una de las alarmas más significativas es la relacionada con la comprobación de la hora del servidor, respecto a una referencia dada. En general en España se utiliza como referencia la hora proporcionada por el Real Instituto y Observatorio de la Armada, aunque puede utilizarse otra.

Por ello, es posible generar una alarma relacionada con el indicador de disponibilidad A2-NTP-HTTP, que detecta una diferencia entre el reloj interno del ordenador, donde se aloja el portal web, y un reloj de referencia, activándose dicha alarma si es superior a un valor umbra, que puede ser por ejemplo de 60 segundos. La secuencia de operaciones es la siguiente:

- obtener la marca horaria que remite el servidor del portal a analizar;
- obtener la hora de referencia; realizar la petición de hora al Real Instituto y Observatorio de la Armada en San Fernando (ROA). La elección de esta entidad es porque es la que establece la hora oficial en España. Sin embargo podría establecerse otra entidad de referencia a solicitud del analista de datos.
- comparan ambas marcas temporales que deben capturarse simultáneamente para poder ser comparables y almacenar la diferencia en el sistema;
- almacenar las diferencias horarias durante el periodo de retención que el analista de datos estime conveniente;
- realizar un estudio de todas las diferencias temporales almacenadas para esa entidad y página web;
- determinar si se ha superado alguno de los umbrales predeterminados;

- ponderar la discrepancia del tiempo diferencial con el oficial para obtener la valoración de la métrica, siendo mayor la puntuación cuanto más cerca del valor oficial esté.

También se puede generar una alarma cuando se detecta un certificado digital con una vigencia inferior a un tiempo predeterminado, como pueden ser 30 días, por ejemplo, siendo la secuencia de operaciones, a realizar las siguientes:

- conectarse vía protocolo https al servidor del portal a analizar;
- extraer la fecha de validez del certificado del servidor;
- comparar dicha fecha con la fecha actual en el sistema y almacenar la diferencia en el sistema;
- almacenar el dato de la diferencia de fechas, durante el periodo de retención que el analista de datos estime conveniente;
- realizar un estudio de todas las diferencias temporales almacenadas para esa entidad y página web;
- determinar si se ha superado alguno de los umbrales predeterminados (menor 30 días);
- ponderar el valor almacenado, siendo mayor la puntuación cuanto más tiempo en el futuro tenga validez el certificado.

Otra posible alarma se puede generar cuando se detecta una velocidad de descarga inferior a un valor preestablecido, para cada grupo de entidades, para ello la secuencia de operaciones a realizar sería la siguiente:

- guardar como marca temporal la fecha y hora del inicio de la petición;
- realizar la conexión a la página web, descargando el contenido;
- guardar como marca temporal la fecha y hora del fin de la petición;
- calcular el tamaño de la petición;
- calcular el tiempo empleado en la petición, como la diferencia entre la marca temporal del fin de la petición y la marca temporal del inicio de la petición;
- calcular la velocidad de descarga como el tamaño de la petición dividido entre el tiempo empleado;
- almacenar la velocidad de descarga junto con la fecha en la que se calculó, tomando valor durante el periodo de retención que el analista de datos estime conveniente;

- realizar un estudio de todas las velocidades de descarga almacenadas para esa entidad y página web;
- determinar si se ha superado alguno de los umbrales predeterminados;
- ponderar el valor almacenado, siendo mayor la puntuación cuanto más velocidad se haya establecido en un momento determinado.

Otra posible alarma se puede generar cuando se detecta que la disponibilidad de una página web es inferior al valor preestablecido para cada grupo de entidades, siendo la secuencia de operaciones a realizar la siguiente:

- guardar como marca temporal la fecha y hora del inicio de la petición;
- realizar la conexión a la página web, descargando el contenido;
- guardar como marca temporal la fecha y hora del fin de la petición;
- si la petición es correcta, se establece ese periodo de tiempo como 100% disponible;
- si la petición es incorrecta, se establece ese periodo de tiempo como indisponible;
- almacenar las indisponibilidades junto con la fecha en la que se produjo, tomando valor durante el periodo de retención que el analista de datos estime conveniente;
- realizar un estudio de todas las indisponibilidades almacenadas para esa entidad y página web;
- determinar si se ha superado alguno de los umbrales predeterminados;
- ponderar los valores almacenados, siendo mayor la puntuación cuanta más disponibilidad tenga la entidad.

Otra posible alarma se genera cuando detecta que el número de cookies de una página web es superior al valor preestablecido para cada grupo de entidades, siendo la secuencia de operaciones necesarias para su detección las siguientes:

- guardar como marca temporal la fecha y hora del inicio de la petición;
- realizar la conexión a la página web, descargando el contenido;
- guardar como marca temporal la fecha y hora del fin de la petición;
- calcular las cookies recibidas en la petición;
- almacenar el número de cookies junto con la fecha en la que se produjo la conexión, tomando valor durante el periodo de retención que el analista de datos estime conveniente;

- determinar si se ha superado alguno de los umbrales predeterminados en número de cookies;
- ponderar los valores almacenados, siendo mayor la puntuación cuanto menos número de cookies se hayan detectado.

Otra posible alarma se puede generar cuando se detecta que el número de enlaces de una página web es superior al valor preestablecido para cada grupo de entidades, siendo la secuencia de operaciones:

- guardar como marca temporal la fecha y hora del inicio de la petición;
- realizar la conexión a la página web, descargando el contenido;
- guardar como marca temporal la fecha y hora del fin de la petición;
- calcular el número de enlaces que la petición ha devuelto;
- almacenar el número de enlaces junto con la fecha en la que se produjo la conexión, tomando valor durante el periodo de retención que el analista de datos estime conveniente;
- determinar si se ha superado alguno de los umbrales predeterminados en número de enlaces;
- ponderar los valores almacenados, siendo mayor la puntuación cuanto menos número de enlaces se hayan detectado.

Inicialmente los rangos utilizados en las distintas alarmas parten de un valor estático, pero puede evolucionar de una manera dinámica, tal que vaya el sistema aprendiendo y detectando dichas alarmas, de acuerdo con el analista de datos y los eventos detectados.

La definición del resto de las alarmas es similar, diferenciándose en el hecho que las desencadena y en el valor del umbral inferior. La superación de los umbrales puede tener alarmas asociadas. Dichas alarmas pueden ser clasificadas según las acciones que se asocien:

- Acciones asociadas a la generación de una alarma para poner en conocimiento de la situación acaecida. Esta alarma puede ser entendida de manera amplia, haciendo actuaciones que permitan notificar a las personas indicadas dicha alarma a través de estímulos sensoriales humanas (vista, oído, tacto, gusto u olfato) como otras que estén vinculadas a la automatización de procesos (p.ej.

modificaciones electromagnéticas). Aunque de manera preferente suele dotarse de alarmas sonoras, iluminación de aviso, dispositivos hápticos de notificación (p.ej. dispositivos con retroalimentación háptica), pero podría extenderse con dispositivos que cambian el entorno a nivel olfativo. A modo de ejemplo podrían ser alarmas dentro de estos ámbitos la remisión de correos electrónicos de alarma, remisión de mensajes vía SMS a dispositivos móviles, mensajes por mensajería instantánea, remisión de información de alerta a redes sociales tipo Facebook/Twitter/Instagram, entre otros.

- Acciones asociadas a la corrección del propio evento en sí, de tal manera que se permita ejecutar acciones que corrijan o al menos mitiguen el propio evento de alerta en sí. A modo de ejemplo aunque no extensivo, se puede indicar que al superar el umbral inferior de validez de un dominio de Internet se realicen las acciones pertinentes para solicitar la renovación automática del dominio DNS.

Debido a la gran cantidad de datos que se irán almacenando de todos los indicadores, se podrá realizar un análisis estadístico para determinar los automatismos para el ajuste de los umbrales anteriormente indicados. Técnicas de segmentación, redes neuronales, clustering, etc pueden determinar un ajuste más fino de las alarmas según la Temática y el grupo a analizar.

La recuperación de los parámetros del servidor y su evaluación se realiza en tiempo real. No obstante, la captura de los resultados de cada indicador, en algunos casos es puntual a lo largo del día y en otras es periódicamente (como por ejemplo el cálculo de la disponibilidad web que es cada 2 minutos comprobando la conexión). Por tanto, si bien la recuperación es en tiempo real, la evaluación depende directamente del algoritmo matemático que a usar.

Las alarmas descritas pretenden dar una perspectiva de las múltiples posibilidades que ofrece la presente invención, siendo su espíritu más amplio al englobar muchos otros posibles indicadores y combinaciones de indicadores para obtener nuevas conclusiones. Hay que tener en cuenta que pueden ir apareciendo test de seguridad que ahora son errores ocultos y que podrán ponderarse tras su descubrimiento, e integrarse en las temáticas descritas o en otras similares.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

Para complementar la descripción que se está realizando y con objeto de ayudar a una mejor comprensión de las características de la invención, se acompaña como parte integrante de dicha descripción, un dibujo en donde con carácter ilustrativo y no limitativo, se ha representado lo siguiente:

La figura 1: Descripción esquemática del procedimiento descrito donde puede apreciarse primeramente como se accede a datos públicos de los diferentes sitios web (1) a evaluar, siendo recogido en el servidor del sistema (2) para su tratamiento. De acuerdo a cada una de las Temáticas (3) que pueden evaluarse, en cada caso se determinan los valores obtenidos en cada Indicador (4) y el valor ponderador (5) en cada caso. Resultado de la interacción de cada uno de los anteriores en cada temática concreta, resulta en la obtención de una Puntuación (6) por temática. La combinación de las diferentes puntuaciones (6) determina el Indicador global CIO-Rank (7) y la generación de los informes y alarmas (8) adecuados en cada caso.

REALIZACIÓN PREFERENTE DE LA INVENCIÓN

La generación de una alarma viene determinada fundamentalmente por cambios inesperados de un indicador de los definidos en la presente patente con respecto a un umbral de referencia. Los umbrales se definen para cada indicador concreto. Para un indicador normalmente se puede establecer un valor o rango de referencia, donde el comportamiento se puede considerar normal. Si dicho comportamiento se desvía de los parámetros de normalidad, entonces pueden suceder dos opciones (en el caso más simple):

- Que el valor actual sea superior el rango de referencia. En este caso se puede establecer una alarma que alerte de niveles superiores al indicado.
- Que el valor actual sea inferior al rango de referencia. En este caso se puede establecer otra alarma que alerte de niveles inferiores a lo que se esperaba.

Este sería el caso de generación de dos alarmas vinculadas a un indicador, pero se pueden definir más alarmas, de tal manera que se puedan establecer múltiples rangos de superación de la normalidad, con comportamientos distintos para cada una de ellas. A efectos comparativos es similar a crear una función matemática definida a trozos, donde cada salto de la función a representar estaría representado por el rango de cada subfunción y su alarma correspondiente. Esto no limitaría el número de alarmas a definir dentro de una métrica o indicador.

En el sistema descrito se identifican alarmas analizando el valor anterior con respecto al valor actual. Si el comportamiento se considera que no está dentro de la normalidad del indicador/métrica, el sistema desencadena el aviso de la situación (por diferentes canales de comunicación, como email, SMS, mensajería instantánea, etc…) y, si es deseable, un comportamiento adicional vinculado a acciones correctivas.

A continuación se describe un ejemplo completo de evaluación, con independencia de las métricas que se capturen:

1. Se carga en el sistema la dirección web del sistema a analizar. Esta carga fundamentalmente requiere la determinación de datos identificativos de la entidad, datos de contacto y datos de servidores que están expuestos al público, para su análisis.

2. Tras la carga de los datos en el sistema, se selecciona qué grupo de entidades es el que le representa mejor, si fuera conocido. Si no lo fuese, se asigna a un grupo general para que comience la captura de la información de cada indicador.

3. Cada indicador se actualiza una vez al día (o varias veces si fuera necesario por la propia necesidad de la métrica, como por ejemplo la disponibilidad web). La actualización se realiza a través de scripts que realizan fundamentalmente cuatro tareas:

- Conexión al servidor destino.
- Extracción de la información
- Tratamiento de la información
- Almacenamiento de los resultados en la plataforma

4. Tras el almacenamiento de los resultados obtenidos, automáticamente se realizan los cálculos de detección de situaciones anómalas. Estas situaciones anómalas están preestablecidas en los propios indicadores a través de disparadores, que detectan los eventos de superación de umbrales. En tales casos, se ejecutan tareas para:

- Notificar mediante las alarmas preestablecidas de la situación encontrada.
- Automatización de las tareas que restauran la situación a la normalidad. En estos casos es necesaria conectividad y acceso a los sistemas finales para poder establecer y ejecutar dichas tareas, o en su defecto, disponer de un frontal de comunicación que permita notificarlo y ejecutar dichas acciones en remoto.

5. Tras la acumulación de resultados y según establezca el analista de datos, se exportará todas las métricas e indicadores, pudiendo ejecutar dichos resultados con programas de análisis estadístico, programas de inteligencia artificial o similares.

6. Tras el análisis de las métricas a través de los diferentes algoritmos que determine el analista de datos, se importarán los resultados en la plataforma, de tal manera que se puedan disponer los valores obtenidos por cada Temática en dicha

entidad analizada. Con los resultados globales por cada Temática, se obtiene el meta-indicador CIO-Rank que consolida los resultados de cada Temática y que sirve como referencia sencilla para los usuarios sin conocimientos técnicos.

7. Con todos los resultados ya calculados, se realiza un informe de la situación de la entidad en base a los resultados obtenidos, siendo el analista de datos el que proporcione explicación de lo obtenido y pudiendo indicar las acciones presentes y futuras para la mejora de los resultados.

REIVINDICACIONES

1.- Método implementado por ordenador para la evaluación del comportamiento de portales web, comprendiendo el método:

- recuperar información técnica del portal a evaluar;
- procesar de los datos recuperados de acuerdo a las condiciones definidas previamente en el propio método;

caracterizado porque el método comprende además la:

- caracterización de las temáticas de indicadores a evaluar;
- caracterización de los indicadores a evaluar dentro de cada temática;
- ponderación de los indicadores seleccionados dentro de cada temática;
- ponderación de cada temática;
- generación de alarmas con el resultado de la evaluación global del portal.

2.- Método implementado por ordenador, según la reivindicación 1, caracterizado porque comprende, al menos, una temática relacionada con la seguridad del portal, una temática con la disponibilidad del portal y una temática del posicionamiento del portal.

3.- Método implementado por ordenador, según cualquiera de las reivindicaciones anteriores, caracterizado porque dentro de la temática de seguridad del portal incluye al menos alguno de los siguientes indicadores:

- certificados HTTPS;
- HTTPS-SHA1;
- número de puertos abiertos;
- Documentos publicados con información institucional privada
- calificación de certificados SSLlabs;
- análisis de robots;
- tipo de servidor;
- seguridad infantil e
- indicador de confianza.

4.- Método implementado por ordenador, según cualquiera de las reivindicaciones anteriores, caracterizado porque dentro de la temática de disponibilidad del portal incluye al menos alguno de los siguientes indicadores:

- velocidad de descarga;
- reloj (NTP-HTTP);
- número de clones web por DNS;
- finalización de la reserva del dominio DNS;
- tiempo de respuesta del servidor
- SLA web
- indicador de lista negra.

5.- Método implementado por ordenador, según cualquiera de las reivindicaciones anteriores, caracterizado porque dentro de la temática de posicionamiento del portal incluye al menos alguno de los siguientes indicadores:

- velocidad de paginado – ordenadores;
- velocidad de paginado – dispositivos móviles;
- ranking tráfico de Alexa;
- número de cookies;
- resultado del buscador Bing;
- resultado del buscador Google y
- número de links.

6.- Método implementado por ordenador, según cualquiera de las reivindicaciones anteriores, caracterizado porque la ponderación de los indicadores, dentro de cada temática, y de las temáticas, se realiza en función de la entidad o grupo de entidades a evaluar.

7.- Método implementado por ordenador, según cualquiera de las reivindicaciones anteriores, caracterizado porque se genera una alarma cuando un indicador de los definidos anteriormente tiene un cambio inesperado respecto a un umbral de referencia.

8.- Método implementado por ordenador, según cualquiera de las reivindicaciones anteriores, caracterizado porque el sistema genera una alarma cuando detecta una

diferencia entre el reloj interno del ordenador, donde se aloja el portal web, y un reloj de referencia superior a un umbral determinado, de acuerdo con la secuencia de operaciones que comprende:

- obtener la marca horaria que remite el servidor del portal a analizar;
- obtener la hora de referencia;
- comparar ambas marcas temporales y almacenar la diferencia en el sistema;
- almacenar las diferencias horarias durante el periodo de retención que el analista de datos estime conveniente;
- realizar un estudio de todas las diferencias temporales almacenadas para esa entidad y página web;
- determinar si se ha superado alguno de los umbrales predeterminados;
- ponderar la discrepancia del tiempo diferencial con el oficial para obtener la valoración de la métrica, siendo mayor la puntuación cuanto más cerca del valor oficial esté.

9.- Método implementado por ordenador, según cualquiera de las reivindicaciones anteriores, caracterizado porque el sistema genera una alarma cuando detecta un certificado digital con una vigencia inferior a tiempo establecido, de acuerdo con la secuencia de operaciones que comprende:

- conectase vía protocolo https al servidor del portal a analizar;
- extraer la fecha de validez del certificado del servidor;
- comparar dicha fecha con la fecha actual en el sistema y almacena la diferencia en el sistema;
- almacenar la diferencia de fechas durante el periodo de retención que el analista de datos estime conveniente;
- realizar un estudio de todas las diferencias temporales almacenadas para esa entidad y página web;
- determinar si se ha superado alguno de los umbrales predeterminados;
- ponderar el valor almacenado, siendo mayor la puntuación cuanto más tiempo en el futuro tenga validez el certificado.

10.- Método implementado por ordenador, según cualquiera de las reivindicaciones anteriores, caracterizado porque el sistema genera una alarma cuando detecta una

velocidad de descarga inferior al preestablecido para cada grupo de entidades, de acuerdo con la secuencia de operaciones que comprende:

- guardar como marca temporal la fecha y hora del inicio de la petición;
- realizar la conexión a la página web, descargando el contenido;
- guardar como marca temporal la fecha y hora del fin de la petición;
- calcular el tamaño de la petición;
- calcular el tiempo empleado en la petición como la diferencia entre la marca temporal del fin de la petición y la marca temporal del inicio de la petición;
- calcular la velocidad de descarga como el tamaño de la petición dividido entre el tiempo empleado;
- almacenar la velocidad de descarga junto con la fecha en la que se calculó, tomando valor durante el periodo de retención que el analista de datos estime conveniente;
- realizar un estudio de todas las velocidades de descarga almacenadas para esa entidad y página web;
- determinar si se ha superado alguno de los umbrales predeterminados;
- ponderar el valor almacenado, siendo mayor la puntuación cuanto más velocidad se haya establecido en un momento determinado.

11.- Método implementado por ordenador, según cualquiera de las reivindicaciones anteriores, caracterizado porque el sistema genera una alarma cuando detecta que la disponibilidad de una página web es inferior al valor preestablecido para cada grupo de entidades, de acuerdo con la secuencia de operaciones que comprende:

- guardar como marca temporal la fecha y hora del inicio de la petición;
- realizar la conexión a la página web, descargando el contenido;
- guardar como marca temporal la fecha y hora del fin de la petición;
- si la petición es correcta, establecer ese periodo de tiempo como disponible;
- si la petición es incorrecta, establecer ese periodo de tiempo como indisponible;
- almacenar las indisponibilidades junto con la fecha en la que se produjo, tomando valor durante el periodo de retención que el analista de datos estime conveniente;
- realizar un estudio de todas las indisponibilidades almacenadas para esa entidad y página web;
- determinar si se ha superado alguno de los umbrales predeterminados;

- ponderar los valores almacenados, siendo mayor la puntuación cuanto más disponibilidad tenga la entidad.

12.- Método implementado por ordenador, según cualquiera de las reivindicaciones anteriores, caracterizado porque el sistema genera una alarma cuando detecta que el número de cookies de una página web es superior al valor preestablecido para cada grupo de entidades, de acuerdo con la secuencia de operaciones que comprende:

- guardar como marca temporal la fecha y hora del inicio de la petición;
- realizar la conexión a la página web, descargando el contenido;
- guardar como marca temporal la fecha y hora del fin de la petición;
- calcular las cookies recibidas en la petición;
- almacenar el número de cookies junto con la fecha en la que se produjo la conexión, tomando valor durante el periodo de retención que el analista de datos estime conveniente;
- determinar si se ha superado alguno de los umbrales predeterminados en número de cookies;
- ponderar los valores almacenados, siendo mayor la puntuación cuanto menos número de cookies se hayan detectado.

13.- Método implementado por ordenador, según cualquiera de las reivindicaciones anteriores, caracterizado porque el sistema genera una alarma cuando detecta que el número de enlaces de una página web es superior al valor preestablecido para cada grupo de entidades, de acuerdo con la secuencia de operaciones que comprende:

- guardar como marca temporal la fecha y hora del inicio de la petición;
- realizar la conexión a la página web, descargando el contenido;
- guardar como marca temporal la fecha y hora del fin de la petición;
- calcular el número de enlaces que la petición ha devuelto;
- almacenar el número de enlaces junto con la fecha en la que se produjo la conexión, tomando valor durante el periodo de retención que el analista de datos estime conveniente;
- determinar si se ha superado alguno de los umbrales predeterminados en número de enlaces;

- ponderar los valores almacenados, siendo mayor la puntuación cuanto menos número de enlaces se hayan detectado.

14.- Método implementado por ordenador, según cualquiera de las reivindicaciones anteriores, caracterizado porque la recuperación de los parámetros del servidor se realiza en tiempo real y su evaluación se realiza en tiempo real o periódicamente dependiendo de la naturaleza del indicador.

15.- Sistema para la evaluación del comportamiento de portales web, comprendiendo el sistema:

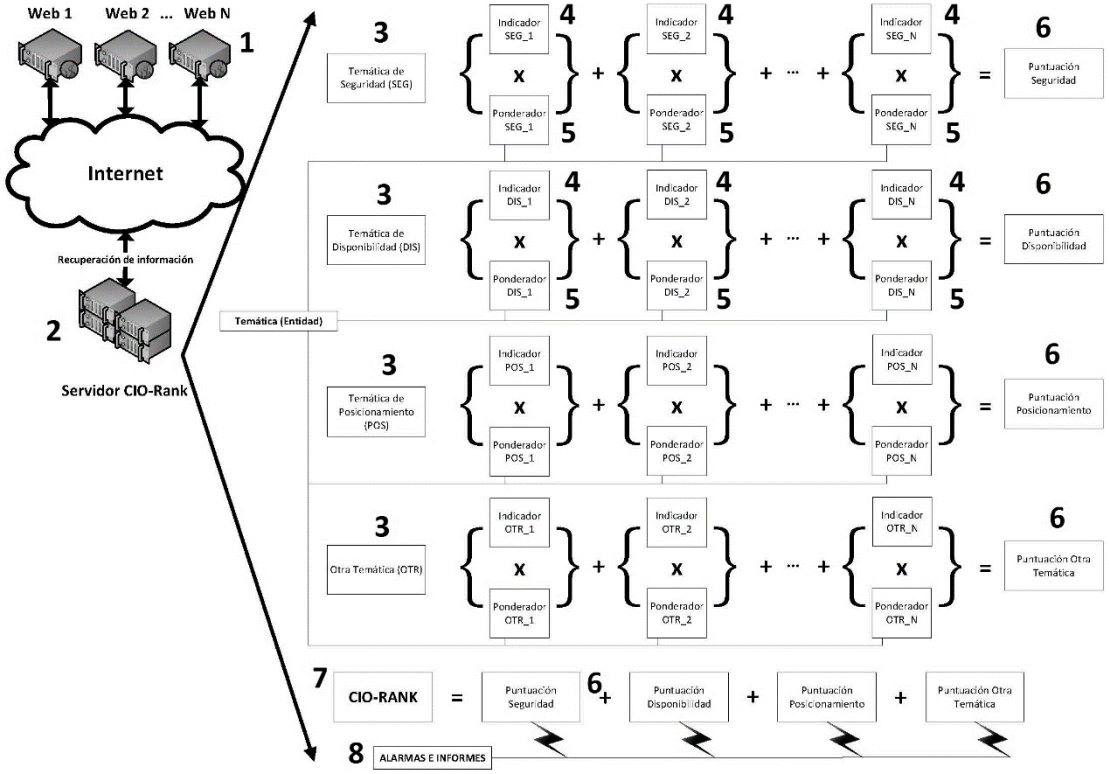
- medios de recuperación y almacenamiento la información técnica del portal a evaluar;
- medios de procesar de los datos recuperados de acuerdo a las condiciones definidas previamente en el propio método

caracterizado porque los medios de procesamiento de datos se configuran además para:

- caracterizar las temáticas de indicadores a evaluar;
- caracterizar los indicadores a evaluar dentro de cada temática;
- ponderar los indicadores seleccionados dentro de cada temática;
- ponderar cada temática;
- generar alarmas con el resultado de la evaluación global del portal.

16.- Producto de programa de ordenador para la evaluación del comportamiento de portales web, caracterizado porque comprende un código de programa que puede usarse en un ordenador para realizar las etapas del método implementado por ordenador definido en cualquiera de las reivindicaciones 1 a 13.

FIG. 1


Web 1
Web 2 ... Web N
1
Internet
Recuperación de información
2
Servidor CIO-Rank
Temática (Entidad)
3
Temática de Seguridad (SEG)
Indicador SEG_1
Ponderador SEG_1
Indicador SEG_2
Ponderador SEG_2
Indicador SEG_N
Ponderador SEG_N
4
5
6
Puntuación Seguridad
3
Temática de Disponibilidad (DIS)
Indicador DIS_1
Ponderador DIS_1
Indicador DIS_2
Ponderador DIS_2
Indicador DIS_N
Ponderador DIS_N
Puntuación Disponibilidad
3
Temática de Posicionamiento (POS)
Indicador POS_1
Ponderador POS_1
Indicador POS_2
Ponderador POS_2
Indicador POS_N
Ponderador POS_N
Puntuación Posicionamiento
3
Otra Temática (OTR)
Indicador OTR_1
Ponderador OTR_1
Indicador OTR_2
Ponderador OTR_2
Indicador OTR_N
Ponderador OTR_N
Puntuación Otra Temática
7
CIO-RANK
Puntuación Seguridad
Puntuación Disponibilidad
Puntuación Posicionamiento
Puntuación Otra Temática
8
ALARMAS E INFORMES

OFICINA ESPAÑOLA
DE PATENTES Y MARCAS

ESPAÑA

(21) N.º solicitud: 201731290

(22) Fecha de presentación de la solicitud: 03.11.2017

(32) Fecha de prioridad:

INFORME SOBRE EL ESTADO DE LA TECNICA

(51) **Int. Cl. :** ***G06Q30/00*** (2012.01)

DOCUMENTOS RELEVANTES

Categoría	(56) **Documentos citados**	**Reivindicaciones afectadas**
X	US 2015039746 A1 (MUKHERJEE RAJATISH et al.) 05/02/2015, Descripción: párs. 21-23	1-16
A	US 2008270209 A1 (MAUSETH MICHAEL JON et al.) 30/10/2008, Descripción: párs. 41-42	1-16
A	US 2017099319 A1 (HUNT ADAM et al.) 06/04/2017, Descripción: párs. 94 y ss.	1-16
A	US 2011270965 A1 (POBLETE BARBARA et al.) 03/11/2011, Todo el documento.	1-16
A	US 2013132213 A1 (LIU BING) 23/05/2013, Todo el documento.	1-16
A	US 6671757 B1 (MULTER DAVID L et al.) 30/12/2003, Todo el documento.	1-16
A	US 2005262240 A1 (DREES TIMOTHY P et al.) 24/11/2005, Todo el documento.	1-16
A	US 2012254405 A1 (GANESH JAI et al.) 04/10/2012, Todo el documento.	1-16

Categoría de los documentos citados
X: de particular relevancia
Y: de particular relevancia combinado con otro/s de la misma categoría
A: refleja el estado de la técnica
O: referido a divulgación no escrita
P: publicado entre la fecha de prioridad y la de presentación de la solicitud
E: documento anterior, pero publicado después de la fecha de presentación de la solicitud

El presente informe ha sido realizado

☒ para todas las reivindicaciones ☐ para las reivindicaciones nº:

Fecha de realización del informe	**Examinador**	**Página**
15.11.2018	M. Muñoz Sanchez	1/2

Documentación mínima buscada (sistema de clasificación seguido de los símbolos de clasificación)

G06Q

Bases de datos electrónicas consultadas durante la búsqueda (nombre de la base de datos y, si es posible, términos de búsqueda utilizados)

INVENES, EPODOC, WPI