

Norma N° 1040/012

Documento original

Ver imagen

- **Fecha de Publicación: 27/02/2014**
- Página: 4
- Carilla: 4

PODER EJECUTIVO

PRESIDENCIA DE LA REPÚBLICA

ÓRGANOS DESCONCENTRADOS

AGENCIA PARA EL DESARROLLO DEL GOBIERNO DE GESTIÓN ELECTRÓNICA Y LA

SOCIEDAD DE LA INFORMACIÓN Y EL CONOCIMIENTO - AGESIC

UNIDAD REGULADORA Y DE CONTROL DE DATOS PERSONALES - URCDP

Resolución 1.040/012

Recomiéndase a los responsables de publicar contenidos en los sitios web de los organismos públicos, la adopción de alguno de los criterios técnicos que se disponen. (160*R)

CONSEJO EJECUTIVO DE LA UNIDAD REGULADORA Y DE CONTROL DE DATOS PERSONALES

RESOLUCION N° EXPEDIENTE N°
1040 2012 2012-2-10-0000634

Montevideo, 20 de diciembre de 2012.

VISTO: La solicitud de informe técnico al CERT-uy, sobre posibles soluciones técnicas para evitar que ciertos contenidos sean indexados e incluidos en el caché de los buscadores de Internet, con posible afectación del derecho a la Protección de Datos Personales.

RESULTANDO: I.- Que toda persona física o jurídica, que publica determinado contenido, tiene que colaborar con su posterior eliminación si ello es solicitado, pues no es posible eliminar un contenido particular de Internet sin la intervención de quien lo publicó.

II.- Que se trata de determinar si desde el punto de vista técnico, es posible evitar que ciertos contenidos, que fueron borrados de Internet, permanezcan indexados y sean mostrados por los buscadores de páginas web.

CONSIDERANDO: I) Que los buscadores periódicamente recorren las páginas web, creando una base de datos y un "caché" con lo que buscan en ellas, de modo de facilitar las próximas búsquedas e indexando además los contenidos (crawling). Esto hace que incluso una vez modificado el sitio real, exista una "copia" del contenido en los sitios de los buscadores.

II) Que cada buscador tiene su propia política de "cache"; por ello en algunos casos es posible advertir que no se desea incluir en el "caché" determinado contenido, mientras que en otros casos hay que solicitar que se elimine posteriormente.

III) Que en todo caso procede la adopción de medidas preventivas, a saber:

a) que los documentos se generen en tiempo real en lugar de que sean generados previamente, de forma que no se puedan almacenar porque simplemente no existen en el momento en que se buscan, controlando así su propagación.

b) que la publicación de información con datos personales se efectúe en formato imagen con etiquetas o atributos que no contengan esos datos.

ATENCIÓN: a lo precedentemente expuesto:

El Consejo Ejecutivo de la Unidad Reguladora y de Control de Datos Personales

RESUELVE:

I - Recomendar a los responsables de publicar contenidos en los sitios web de los organismos públicos, la adopción de alguno de los criterios técnicos señalados en la parte expositiva de este acto, que se desarrollan en el informe técnico del CERT-uy, que se incorpora a esta resolución.

II - Comuníquese, publíquese y oportunamente, archívese.

Fdo.: Dr. Felipe Rotondo, Consejo Ejecutivo URCDP.

Informe técnico Expediente N°: 2012-2-10-0000634

Introducción

A raíz de la solicitud realizada por la Unidad Reguladora y de Control de Protección de Datos Personales, se realizó el análisis de opciones para poder eliminar un contenido de los buscadores de Internet una vez que este fue indexado y cacheado.

En todos los casos, la persona que publica el contenido tiene que colaborar con su eliminación, no es posible eliminar un contenido particular de Internet sin la colaboración (aunque bajo obligación) de la persona que subió el contenido. Un ejemplo claro de eso son los sitios de contenidos multimedia piratas.

El objetivo es entonces determinar si es posible evitar que contenidos que fueron borrados en Internet permanezcan indexados y mostrados en los caches de los buscadores de Internet como ser Google, Bing o Yahoo.

Contexto

Los buscadores de Internet (Google, Bing, Yahoo, etc) periódicamente recorren Internet indexando los contenidos (crawling). De esta forma crean una base de datos para realizar las búsquedas de contenidos en Internet.

Al mismo tiempo se crea un caché de los contenidos, lo que hace que incluso una vez modificado el sitio real, exista una "copia" con parte del contenido anterior en los sitios de los buscadores.

Esto ocurre para contenidos de texto, no para imágenes o videos ya que en estos últimos solo se indexan las etiquetas (tags) de los mismos. Es entonces importante tener en cuenta que si se quiere aparecer en los buscadores, el contenido tiene que tener texto, ya sea por ser un documento, o porque existen etiquetas de texto que describen a un contenido multimedia.

Agregar las etiquetas manualmente requiere de un esfuerzo extra, ya que hay que hacerlo de forma tal que esas etiquetas sean útiles para describir el contenido del contenido como ser: "resolución", nombre de la empresa, nombres de los firmantes, título de la resolución, etc.

Análisis

En el escenario antes descrito es importante destacar los siguientes puntos:

- Es necesario ser indexado por los buscadores de Internet en primer lugar para aparecer en los resultados de las búsquedas.
- Si el contenido no es texto, es necesario incorporar etiquetas al contenido de forma manual.
- Cada buscador tiene su propia política de caché, donde en algunos casos es posible advertir que no se quiere cachear el contenido, mientras que en otros casos hay que solicitar que se elimine posteriormente. Incluso en algunos casos no hay forma de prevenir o eliminar el caché.
- Mas allá de los buscadores, un usuario siempre puede copiar el contenido a otro sitio, y en este caso el buscador lo encontrará e indexará en el otro sitio.

Con este escenario, la opción que más aplica para evitar ser cacheado, y que una vez que el contenido se borre no aparezca más en los buscadores, es que los documentos se generen en tiempo real a raíz de peticiones en lugar de que sean documentos generados previamente. De esta forma los buscadores no podrán cachear el contenido porque simplemente no existe en el momento en que consultan. Sin embargo, esto hace necesario que se incluyan etiquetas en los contenidos para que puedan ser indexados y encontrados por los buscadores.

Esta opción sin embargo tiene la desventaja de que no es una solución estándar, y además es necesario agregar las etiquetas manualmente para ser indexados, tarea que no siempre es sencilla y/o viable.

Conclusión

Dada la complejidad, amplitud y masificación de usuarios de Internet, una vez que se sube un contenido, se puede considerar que ya no es posible borrarlo completamente. Para mantener posibilidades de borrar un cierto contenido realmente se tiene que primero haber controlado en cierta forma su copiado y propagación. Para esto existen alternativas que generan el contenido en tiempo real o lo sirven en base a consultas específicas, haciendo que su copiado o indexación sea más compleja, y controlando así su propagación.

El espíritu de los buscadores es que fácilmente se encuentren los contenidos, haciendo la tarea más fácil tanto para quien sube el contenido como para el que lo consume. Al mismo tiempo, los buscadores más populares cuentan con mecanismos para remover los contenidos de los caches. En caso de que ingrese una solicitud de remoción de información del caché de cierto buscador, y que este constata que el sitio original ya no existe, la remoción es prácticamente inmediata. Si la propagación fue controlada en forma adecuada, remover el contenido del sitio original y solicitar la remoción de los cachés de los buscadores puede proveer una posibilidad real de eliminación.

Teniendo en cuenta estos motivos entendemos que la mejor opción es tomar medidas preventivas para controlar la propagación de los documentos, y una vez que se quieren borrar, simplemente borrar el contenido incorrecto de Internet y solicitar a los buscadores más populares que lo remuevan de sus cachés. De esta forma se reduce el riesgo de que se encuentre dicho contenido en Internet, quedando como única posibilidad que un usuario lo haya copiado y subido en otro sitio en forma previa. A continuación se presentan los links para solicitar la baja de los buscadores más populares:

Google

<http://support.google.com/webmasters/bin/answer.py?hl=en&answer=1663660>

Bing

<http://www.bing.com/webmaster/help/block-urls-from-bing-264e560a>

Yahoo

<http://help.yahoo.com/l/us/yahoo/privacy/access/indexing-13.html>

En todos los casos existe también la posibilidad de utilizar metatags en el código de la página u otros artefactos para evitar ser indexado o cacheado, sin embargo las referencias a estos métodos muestran una tasa de fallo muy grande, por lo que no se recomienda confiar en ellas.