FACTURACIÓN CONTABILIDAD

LAS BIBLIOTECAS ENFOCAN SU LABOR EN LACONSERVACIÓN DEL CONOCIMIENTO CREADO EN INTERNET

Una cruzada contra los agujeros negros de internet

TAGS > INTERNET

El depósito legal electrónico británico se aprobó el pasado mes de abril. (EFE)

Verónica Ramírez 10/07/2013 (06:00)

A | A

0 Enviar 19 9 7 2 Menéame Imprimir

Año 2150. El departamento de Historia de una universidad cualquiera se encarga de hacer un análisis político de, por ejemplo, los partidos que protagonizaron las elecciones generales de 2011 en España. Tienen referencias de la existencia de un llamado Partido Pirata. Hacen una búsqueda en internet y *voilà*! No hay información de primera mano. Es decir, no se sabe qué contaba la web original. Lo mismo pasaría con el resto de partidos, sucede con millones de páginas, y no digamos ya con los **hilos de conversación de las redes sociales**. Hay un agujero negro en la historia de España. Pero si en internet estaba todo, ¿no es así?

En absoluto. De hecho, la información digital se destruye en gran proporción a la que se crea. Hay millones de direcciones de páginas web, aquellas siglas conocidas como URL, que ya no existen. Aunque suene exagerado, una página dura de media menos -100 días- que lo que tarda en descomponerse el papel al aire libre -un año-. Parece claro, por tanto, que hay que ponerse al día en la **conservación del patrimonio digital** antes de que se desintegre en la red.

En eso están trabajando instituciones como la Biblioteca Nacional de España (BNE), que ha reunido en su sede madrileña a expertos en la materia de diferentes países en la *Jornada Internacional sobre Archivos Web y Depósito Legal Electrónico.* Con el lema *El patrimonio documental en internet*, responsables de los **principales archivos web de diferentes países** han puesto en común el panorama internacional y el gran reto al que se enfrentan las bibliotecas del mundo: la conservación de los documentos creados en internet.

Aquel departamento de historia del siglo XXII tendrá que recurrir a lugares como *Internet Archive* para encontrar la información original que presentaba aquel partido político en octubre de 2011. La BNE está finalizando en la actualidad **la octava recolección masiva del dominio .es** con la colaboración de esta institución pionera en la búsqueda de páginas y sitios web, que recoge páginas de todo el mundo desde 1996.

Esta recolección, que estará disponible en otoño en la BNE según explicó **Mar Pérez Morillo**, jefa del servicio web de la institución, cuenta con una listado de más de 1,6 millones de nombres de dominios de internet bajo el código de país *.es*. El Archivo de la web española reúne 100 *Terabytes* (TB) de información comprimida en formato WARC (*Web ARChive file format*, el estándar internacional ISO 28500).

Mucha información

La tarea no es fácil. La jornada estuvo protagonizada por el término "retos" de forma equivalente a "archivo digital" o *crawler* - robot de rastreo de páginas-. **Kristine Hanna**, jefa del servicio de archivado de Internet Archive, esbozó los problemas de base a los que debe responder una biblioteca digital: cuánto archivar, cómo hacer la selección, cómo gestionar los documentos… Es decir, **cómo hacer frente a la cantidad ingente de información** que puebla la red.

La institución gestiona, desde 1996, 400.000 horas de televisión, 500.000 imágenes,

MAPFRE, personas que cuidan de personas
MAPFRE

AHORA EN PORTADA

Bruselas declara ilegales las ayudas a los astilleros entre 2007 y 2011

Los rumores de una rebaja del 'rating' a España ponen en alerta a los mercados

El fiscal pide que Bárcenas siga en prisión porque no se puede controlar sus cuentas

CONÉCTATE A LA ACTUALIDAD

SUSCRÍBETE AL BOLETÍN
Boletín Urgentes
Escribe tu email

Seguir Me gusta

LAS MÁS LEÍDAS cajamar
CULTURA EL CONFIDENCIAL

1.- González Vigil pide el cese de
2.- Un nido de serpientes en la SGAE
3.- Arturo Pérez Reverte paga los
4.- El laberinto picassiano a través
5.- "Si el Gobierno no cuida
6.- 'Pacific Rim' se da
7.- Instrucciones para ser mujer
8.- Así se tumba a un presidente
9.- Frikis del mundo, uníos: llega
10.- La tragedia de la atleta olímpica

Web2PDF
converted by Web2PDFConvert.com

950.000 grabaciones sonoras, 2.700.000 libros y **150.000.000.000 de páginas web.** Para Hanna, tan solo un goteo en comparación con el inmenso chorro de datos que expulsa la "boca de riego" que es internet. *La web es un auténtico lío* era el título de su ponencia.

> *Internet Archive gestiona 150.000.000.000 de páginas web*

Más barullo todavía son las redes sociales: "Necesitamos nuevas herramientas para captar estos medios", expresó Hanna, imprescindibles en el futuro para captar formas de expresión de la sociedad que solamente están en la red. *Internet Archive* recopila también **capturas de los perfiles públicos** de las redes sociales.

En la jornada se dieron a conocer multitud de proyectos similares de todo el mundo, como el de la Stanford University de Palo Alto (California), que mantiene entre su colección una específica sobre sitios web y blogs del Medio Oriente desde 2009. Gracias a esa recopilación permanecen almacenada la documentación digital acerca de eventos históricos tan trascendentales como las protestas que tuvieron lugar tras las elecciones de ese año. Entre ese arsenal de información se encuentra la cuenta de Flickr del propio Mir-Hosein Musaví, el político iraní de la oposición, que perdió dichas elecciones y recogió una gran cantidad de imágenes de aquellas manifestaciones.

Una red de archivos web

La **colaboración entre distintas instituciones** fue otra de las premisas que compartieron los ponentes. Una forma de trabajo ineludible a tener en cuenta para poder afrontar los mencionados retos del futuro.

La organización que pone de manifiesto ese principio es el Consorcio Internacional de Conservación de Internet (IIPC), el mayor archivo digital del mundo, fundado en 2003 y que cuenta con 44 instituciones asociadas –entre ellas la BNE-. **Mary Pitt**, responsable del programa y comunicaciones del IIPC, presentó la institución y compartió más ejemplos de espacios de recolección de archivos digitales, diferenciando por su acceso público (archivos claros) o restringido (archivos oscuros):

- Pandora, la biblioteca digital de Australia, uno de los proyectos pioneros en archivo digital, de acceso público.
- Helvetica, el archivo web de Suiza de acceso restringido. Solo se puede acceder estando en la biblioteca.
- *Archives de l'internet*, de la Biblioteca Nacional de Francia (BNF), que cuenta con 20.000 web recolectadas, 370 TB de archivos y con previsión de crecimiento de 100 TB al año.
- British Library: no está accesible, excepto con permiso.

"La conservación digital no es diferente de la tradicional", comentó **Sara Aubry**, responsable técnica del archivo web de la BNF, también presente en la jornada. Aubry explicó con detalle el proceso documental que sigue la recolección: **selección, indexación, conservación, interfaz web** para los usuarios... Cada una de las fases está apoyada por diferentes herramientas web. La documentalista y diseñadora web destacó la labor continua de mejora de la biblioteca francesa, cuyos retos más importantes son una mayor vinculación los archivos con el catálogo de la biblioteca tradicional y lograr una búsqueda inteligente de URL –ahora requiere insertar la URL exacta-.

Las metas que persigue el archivo británico, que cuenta con 14.000 páginas web, es ofrecer los contenidos abiertos al público, ya que ahora se encuentran restringidos, según explicó su responsable **Helen Hockx-Yu**.

País Vasco y Cataluña, pioneras

Muchos datos, mucha información y muchas preguntas. **Icíar Muguerza**, miembro del servicio web de la Biblioteca Nacional de España, planteó la pregunta de dónde establecer los límites de lo prescindible: ¿Dónde se sitúa el punto entre lo que nos interesa y lo que no? Cada uno de los asistentes pusieron de manifiesto esta como una de las mayores dificultades a la que hacer frente. Hasta ahora, cada uno de los archivos está compuesto de **"consejos" formados por profesionales que definen los criterios**.

Por falta recursos o por decisión propia, confían en sus bibliotecarios y en el resto de perfiles que trabajan en los archivos. En el archivo web español se constituyó un "**Consejo de regulación bibliotecaria**", dentro del cual hay un equipo dedicado al depósito legal digital.

Participaron también en la jornada representantes de bibliotecas del País Vasco (2007) y Cataluña (2005) **pioneras del archivo web en España**, además de Telefónica y Red.es, con quien la BNE firmó recientemente un convenio para la construcción de un repositorio común que dé cabida al Depósito Legal Electrónico (DL-e).

El secretario de Estado de Cultura, **José María Lassalle**, inauguró la sesión asegurando asegurado que el decreto que regulará el DL-e, y que ayudará a la preservación del patrimonio cultural, estará vigente "en unos pocos meses", después

Web2PDF
converted by Web2PDFConvert.com

de haber pasado ya los trámites de los organismos consultivos.

Francia y Reino Unido ya cuentan con el DL-e desde 2006 y abril de 2013, respectivamente, tal y como explicaron las representantes de ambas bibliotecas.

El Secretario de Estado destacó la "naturaleza efímera de los recursos que hoy compartimos" y recordó las últimas estimaciones que sitúan la vida media de una página web entre 44 y 75 días de duración (Kristine Hanna amplió esa cifra a 100). Esa escasa duración convierte este asunto en "**uno de los objetivos principales y más urgentes de la Biblioteca Nacional**".

Lassalle mantuvo un tono ciertamente alarmista en torno a este asunto en su intervención, secundado de alguna forma por las ponencias posteriores. El riesgo de la desaparición documental acecha, o eso parece. La conclusión general es que el volumen de información que contiene la red es tal que **"no se puede recuperar todo"**. Aun así, hay que trabajar por conservar lo más posible. Para que las generaciones venideras, ya sea dentro de diez años como en el 2150, sepan con exactitud qué se publicaba en internet un 10 de julio de 2013.

- conoce cuántos intereses pagas en una hipoteca de 15, 20 o 30 añ
- los 27 pisos nuevos con piscina más baratos de españa

0 19 9 7 2 Menéame

EL REDACTOR RECOMIENDA

La Biblioteca Nacional entra en la era digital

El Círculo de Bellas Artes sale a la calle a pedir para salvarse

LA OPINIÓN DE LOS LECTORES

0 Escribe el tuyo

DEJA TU COMENTARIO | VER MÁS COMENTARIOS

BOLETÍN
HEMEROTECA

España
Economía
Comunicación
Vivienda
Sociedad
Mundo

A. Casado
C. Sánchez
J.A. Zarzalejos
F. Quevedo
Otros columnistas
Interiores
El Confidente

Al minuto
Mercados
Cotizaciones
Fondos
Finanzas Personales

Emprendedores
Techfacts
Siliconvalleyfa.
Metaemprended.
Directivo Digital

Liga BBVA
NBA
Fútbol
Fórmula 1
Baloncesto
Tenis
Golf
Atletismo

Bienestar
Ciencia y Salud
Educación
Sociedad
Sexualidad
Blogs

Libros
Arte
Música
Cine
Escena

Corazón
Casas Reales
Televisión
Cine
Moda
Belleza
Ocio

Vídeos
Álbumes
Fotos
España is not Spain

Encuentros Digitales
Foro
Editar Perfil
Foro Interno

Copyright 2010 | Quiénes somos | Aviso Legal | Política de Privacidad | **Publicidad** | Contacto | Favoritos | Página de Inicio | Ayuda | RSS | Audiencia medida por Comscore

Web2PDF
converted by Web2PDFConvert.com