

REFLEXIONES SOBRE EL CICLO DE VIDA DE LOS SITIOS WEB COMO FUENTES DE INFORMACIÓN EN PROCESOS DE INVESTIGACIÓN Y TRABAJOS ACADÉMICOS

José Pablo Meza Pérez *
jmeza@cariri.ucr.cr.ac.cr

Fecha de recepción: 27 de junio 05 / Fecha de aceptación: 20 de setiembre 05

Resumen

Internet se ha posicionado como un recurso importante de comunicación del pensamiento e ideas. Investigadores, académicos y centros. Se debaten sobre el nuevo paradigma en el mundo de la información. Las características de los sitios Web como un documento casi único, sin múltiples copias y de acceso universal. Impone el reto de lograr establecer parámetros de validación de información para ser citado como una fuente de información confiable en procesos de investigación y sustentar con bases sólidas y perdurables el nuevo conocimiento.

El proceso de investigación de un grupo de sitios Web por un periodo de 15 meses, muestra resultados importantes para conocer esa nueva realidad informativa. En el transcurso de este periodo, el total de sitios Web que no mostraron actividad y reportaron modificaciones en su contenido y forma, represento cerca del 42 por ciento de la muestra. Solo el 58 por ciento de los sitios Web fue igual a los utilizados por sus autores para la elaboración de sus trabajos, pese al hecho de hacer un balance aceptable de los recursos de información.

Palabras clave: investigación, internet, ciclo de vida, bibliometría, fuentes de investigación, costa rica, comunicación, sitios web.

Abstract

Internet has taken a stand as an important resource for communication of ideas and thought. Scholars, research centers and professionals argue about this new information paradigm. The web sites represent a unique type of document that can be accessed globally although only one copy is available. Therefore, it is a challenge to determine the information validation standards that tell us which sources are to be trusted when conducting quality research. The new knowledge could then be supported under reliable basis.

The present paper shows the results of this new information reality through a research, conducted among several web sites, during a period of fifteen months. During that period, 42% of the sites did not show any activity neither changed content or design. Only 58% of the sites were used by authors to write their papers, despite the fact that there were a large variety of information resources that met quality standards.

Keywords: information research; web sites; information sources; Costa Rica.

* Profesor de la Carrera de Archivística, Universidad de Costa Rica.

Introducción

Internet se suma hoy como otro recurso más de información en los procesos de investigación y trabajos académicos. Es un recurso de amplia difusión. Sin embargo esta fuente de información posee características particulares que hace necesario establecer criterios metodológicos rigurosos para su selección como fuente de información válida y confiable.

Las tecnologías de información imponen un nuevo paradigma en los procesos de investigación en el ámbito académico y en la forma de conceptuar el mundo de la información. La proliferación de fuentes de información ha favorecido un verdadero crecimiento exponencial de los recursos de investigación y comunicación del pensamiento.

Esta red de comunicaciones desde su origen como ARPANET en la década de los años 70 (Abaitua, J., 1999), se ha logrado posicionar como una herramienta alternativa e importante para la investigación, rompiendo las barreras de espacio y del tiempo para la transmisión de ideas. Se impone hoy el reto de impulsar la alfabetización informativa para la adquisición de habilidades en el manejo de las tecnologías de información y acortar la brecha digital.

El uso que se da a Internet en procesos de investigación, paradójicamente se ha convertido en la capacidad de acceso a nuevos recursos de información casi en forma inmediata, pero también presenta una amenaza por el cambio vertiginoso de esta fuente de información. Su versatilidad de comunicación es proporcional a la capacidad o velocidad de perder información relevante y oportuna para la ciencia.

La falta de guía oportuna de los educadores o formadores académicos, ha promovido que un buen número de trabajos académicos muestre la tendencia a citar este recurso en forma importante, excluyendo otras fuentes de información que no se encuentran en soporte electrónico. Abandonando los principios básicos de la metodología científica como: exhaustividad en la consulta de fuentes de información, la verificación de datos, la consulta a fuentes primarias de información, la validación de información y otros (Bunge. 1985 p.41).

Las implicaciones metodológicas y científicas que acarrea el ciclo de vida de las páginas Web, puede dejar sin sustento un proceso de investigación en un segundo y es una amenaza real para validar los aportes de la investigación a la formación del conocimiento y resta credibilidad, rigurosidad, entre otros aspectos a este proceso.

En las publicaciones impresas se encuentra establecido su control universal, basa su principio en depositar toda producción en sistemas nacionales de bibliotecas, archivos y centros de investigación.

Al no existir un mecanismo de aplicación general para la salvaguarda de las publicaciones electrónicas en la red que garantice el derecho a la información, es desconocida la magnitud del problema con relación a la pérdida de información científico - cultural en Internet. Existen esfuerzos en ese sentido; las bibliotecas están jugando un papel importante en lograr formar colecciones digitales que reúnan alguna parte de la producción relevante en soporte electrónico para su preservación en el tiempo y garantizar el acceso a la información. Oportunidad para dimensionar a las bibliotecas en un nuevo rol y ámbito de acción. El proyecto Internet II a mediano o largo plazo parece ser el instrumento para conservar la información con valor científico-cultural e histórica en este soporte electrónico.

En el ámbito de la investigación y como evidencia de la magnitud e importancia de establecer criterios de selección de fuentes de información en Internet, en el artículo

Going, Going, Gone: Lost Internet reference publicado en la revista Science, se reseña un estudio comparativo de tres importantes revistas de carácter científico (Science, New England Journal of Medicine y Journal of the American Medical Association) mencionando, que el promedio de citación de referencias de Internet, no sobrepasa el 2.49 por ciento citas por artículo publicado. En 15 meses el 10 por ciento de estas citas presentaban inactividad y en 24 meses el porcentaje fue del 13 por ciento, dato significativo para ser estudiado (Dellavale, R.P. *et. al*, 31 de octubre 2003: p.787). El mismo hace notar que en el proceso de control editorial de centros de excelencia en la investigación;

la citación de fuentes de Internet pasa por un fuerte escrutinio antes de ser incorporado en una publicación de calidad. Pese a ello los factores de variabilidad de la red se hacen presentes y comprometen las bases que sustentan la formación del nuevo conocimiento.

La observación de este comportamiento, hace suponer las dificultades que existen pero son pocos los estudios al respecto. Generar datos relevantes que permitan conocer, validar la consistencia de esta fuente de información, como elementos para sustentar las premisas para la formación de criterios académicos y mejorar los procesos de investigación es una tarea importante. Realizar estudios desde una vivencia cotidiana y pragmática con un enfoque multidisciplinario es tarea importante a desarrollar y urgente.

Aporte de las ciencias de información al estudio de fuentes de información en procesos de investigación

Las ciencias de la información entendida en su forma más genérica su objeto y sujeto de estudio es la información (Goñi Camejo, 2000), y no solo concurren disciplinas informáticas, engloba también a las ciencias bibliotecológicas, documentales, la archivología y otras. Han proliferado nuevas disciplinas y técnicas que se encargan de estudiar fenómenos específicos atinentes a estas ciencias; algunas incorporan a la ciencia métrica al fenómeno información, facilitando el desarrollo y difusión de nuevas interdisciplinas y multidisciplinas como la bibliometría, la infometría y la cibermetría que se encargan de estudiar las características del comportamiento de la información en sus diferentes soportes documentales (Araújo Ruiz y Arencibia 2002). Utilizando técnicas como la citación de referencias, el análisis cuantitativo, formulación de leyes y otras (Urbizagástegui Alvarado, 1999).

Poco son los estudios realizados en Costa Rica. Se destacan los desarrollados en el campo de bibliometría en trabajos finales de graduación en bibliotecología en las Universidades Públicas. También el esfuerzo por mantener la calidad de la investigación hace necesario la normativa y tomar con responsabilidad las implicaciones de este fenómeno. De la discusión emprendida en

la Carrera de Archivística de la Universidad de Costa Rica por normar la calidad de los proyectos y requisitos de graduación del programa de licenciatura nació la iniciativa y motivación para realizar este proceso de investigación y conocer sus repercusiones en el ámbito académico. Como parámetro de calidad de una proyecto de graduación de una Universidad Pública.

Son procesos de investigación que nos permiten darnos luz y aportar sobre aspectos importantes de las características de la citación en procesos de investigación y académicos. Llegar a conocer estos atributos permite a las bibliotecas mejorar su acervo documental y en el plano docente a la formación de programas de capacitación que desarrollen competencias laborales y académicas significativas en la sociedad de la información y el conocimiento.

Sujeto y objeto de estudio

La Unidad de Referencia de la Biblioteca W. K. Kellogg de la Universidad EARTH y el titular de este artículo quien coordinó dicha unidad por un largo periodo. Realizó un proceso de generación de datos métricos para reunir información sobre las características de citación de fuentes de información en el requisito de graduación académica y conocer los recursos de información más utilizados.

Los datos generados después de 15 meses de estudio de dos muestras seleccionadas de proyectos de graduación y tomando de ellos la citación de sitios WEB permiten realizar una aproximación, para conocer el comportamiento y formular criterios de citación de fuentes de información, mejorar así las habilidades informativas de los estudiantes, aportar elementos de calidad académica para la Universidad y validar información para su aplicación en la investigación.

Descripción metodológica y procesamiento de datos

Se escogió dos muestras aleatorias de 32 Proyectos de Graduación de la promoción 2003 y 2004 de la Universidad EARTH, con un nivel de confiabilidad de 95 por ciento y un error no mayor o menor del 5 por ciento.

Las citas bibliográficas fueron incluidas a una base de datos en el programa cds/isis software especializado, desarrollado y distribuido por UNESCO. El insumo primario que se utilizó fue la copia electrónica del documento original entregado a la Biblioteca W. K. Kellogg por parte de los autores.

La migración de citas bibliográficas por proyecto de graduación fue realizada con una herramienta que genera un archivo en códigos ISO norma 2709 (programa texiso) el cual permite importar íntegramente los datos a una base de cds/isis. Se utilizó tablas y gráficos del programa excel y el programa automático de verificación de link "WEB LINK VALIDATOR" generando datos de seis reportes de accesibilidad, fechas de creación y modificaciones de sitios WEB, mediciones realizadas cada cuatro meses iniciando el día 14 de enero del 2004 a la fecha 18 de abril del 2005.

Resultados del procesamiento de datos

Caracterización de las muestras

La muestra A. Se fue integrada con la población 902 citas bibliográficas de 32 proyectos de graduación del año 2003. Estableciendo un valor medio de 24 citas y un rango 15 a

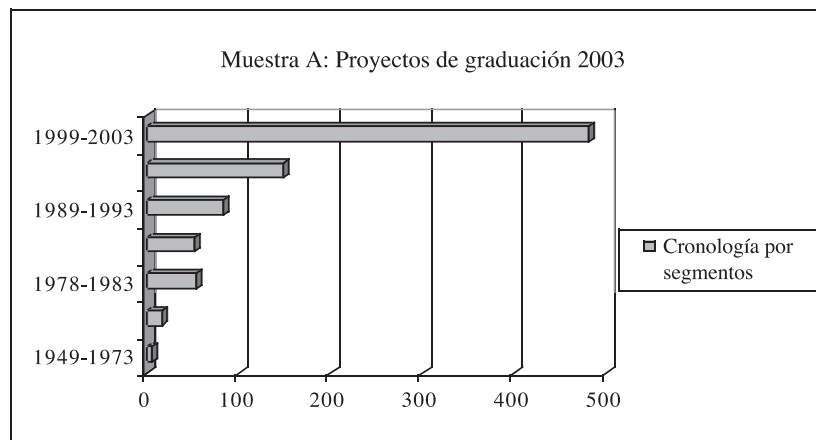
28 referencias por proyectos. Solo se trabajo la bibliografía citada en el cuerpo del texto utilizada por los autores. El total de citas de Internet que se establecieron como activa en el transcurso de los 15 meses de estudio fue 313 direcciones de Internet.

La muestra B. Se integro con la población 844 citas bibliográficas de 32 proyectos de graduación del año 2004. Estableciendo un valor medio 26 citas y un rango de 14 a 38 referencias por proyecto. El total de sitios de Internet que han mostrado actividad de enero al mes de abril del 2005 es de 241. Se han realizado a la fecha dos verificaciones de accesibilidad y se espera realizar la toma física de datos al menos 12 meses más. La muestra B., se utiliza como parámetro de comparación para establecer semejanzas y diferencia con la muestra A., pero no se incluyen los datos de accesibilidad por no ser suficientes.

Resultados

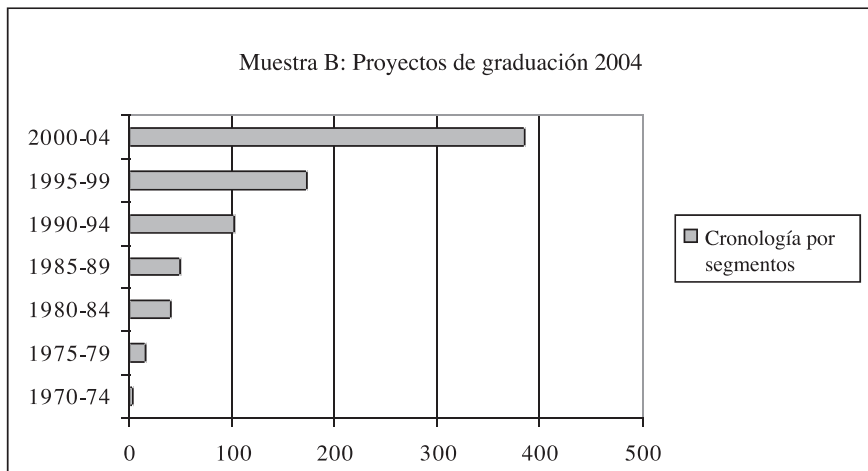
Los datos obtenidos sobre el comportamiento de la citación en los proyectos de graduación. Establece que la mayor citación de fuentes de información por periodos cronológicos la ubica en el rango de los 15 a 20 años de antigüedad (ver gráficos 1 y 2).

Gráfico # 1



Fuente: Datos métricos, proyectos de graduación 2003.

Gráfico # 2



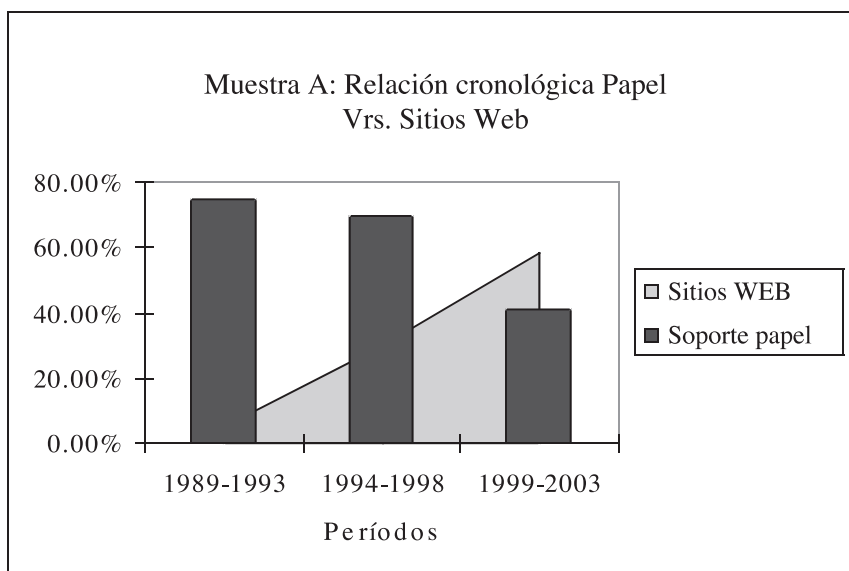
Fuente: Datos métricos proyectos de graduación 2004.

Se muestra una tendencia a consultar fuentes de información recientes, no mayores a 5 años de antigüedad. Valida que no son excluidas otras fuentes de información anteriores a esta fecha.

La relación de fuentes de información cronológicas con las categorías soporte papel y sitios

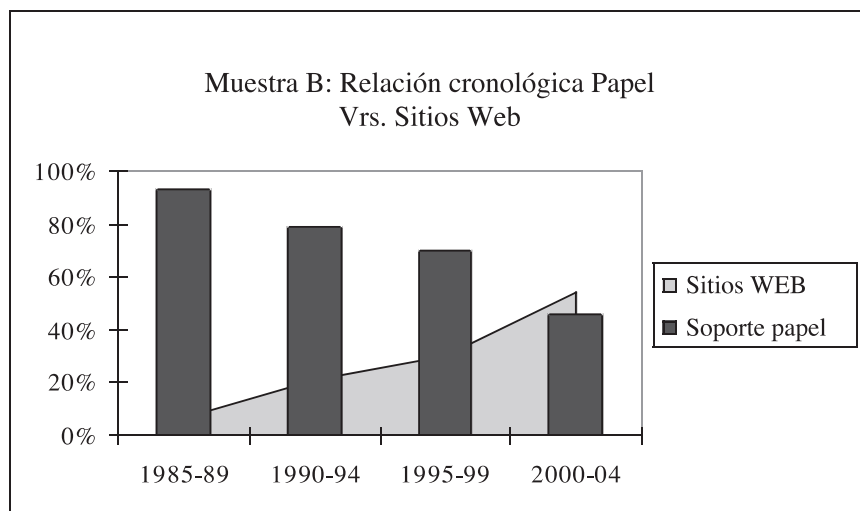
Web. Presenta una tendencia en ambas muestras a ubicar a la categoría sitios WEB como recurso de primero orden para dar actualidad a los proyectos de graduación.

Gráfico # 3



Fuente: Datos métricos proyectos de graduación 2003.

Gráfico # 4



Fuente: Datos métricos proyectos de graduación 2004.

Se destaca con relación a la muestra B. El segmento de citación de información en el periodo 2000 al 2004 o más actual, los valores son aproximadamente iguales en las categorías soporte papel y sitios Web. La categoría de sitios Web solo sobrepasa a la categoría Papel por un 10%. La explicación a este fenómeno se encuentra en el mayor control de calidad por parte de los directores de proyectos de graduación y al esfuerzo de la Unidad de Referencia por presentar avances de resultados del estudio métrico a la generación de los proyectos de graduación 2004.

Un aspecto destacable sobre la citación de sitios WEB es que su comportamiento se invierte en el segmento anterior al más actual y sigue su descenso paulatinamente en contraposición a la categoría de información en soporte papel que representa un desarrollo significativo mayor a los valores del segmento mas actual (ver gráficos 3 y 4).

Categorización de dominios de los sitios Web contenidos en las muestras estudiadas

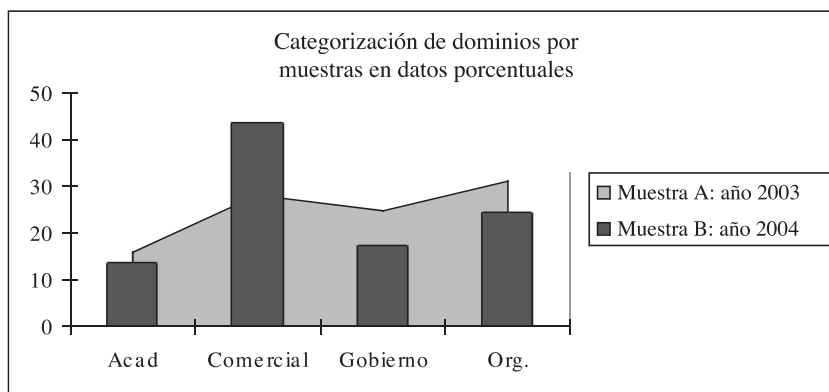
La caracterización de dominios de sitios Web contenidos en ambas muestras se logro agrupar en categorías muy concretas. Los sitios

Web de organizaciones o dominios org; en los cuales se ubicaron todas aquellas organizaciones nacionales, multinacionales y organización no gubernamentales. Otra categoría fue integrada por los sitios Web de vinculación estatal conocidos como los dominios gov. Los sitios Web comerciales o dominios com. Por último los sitios de Web de instituciones académicas o dominios aca o edu.

La utilización de fuentes de información de sitios Web por dominios se resalta una tendencia a citar en mayor número los sitios comerciales, aumentando particularmente el la muestra B. con un 45% y una tendencia a las disminuciones de las otras categoría (ver gráfico 5). El componente empresarial del enfoque de la formación de la Universidad, puede ser un elemento importante sobre la tendencia a citar en mayor número, este dominio.

Sin embargo se debe esclarecer las características particulares de cada uno de ellos según función social, económica, política e ideológica que representan. Debe ser un contenido obligatorio en los procesos de formación de habilidades informativas, cursos de investigación y programas de formación de usuarios de la información.

Gráfico # 5



Fuente: Datos métricos proyectos de graduación 2003-2004.

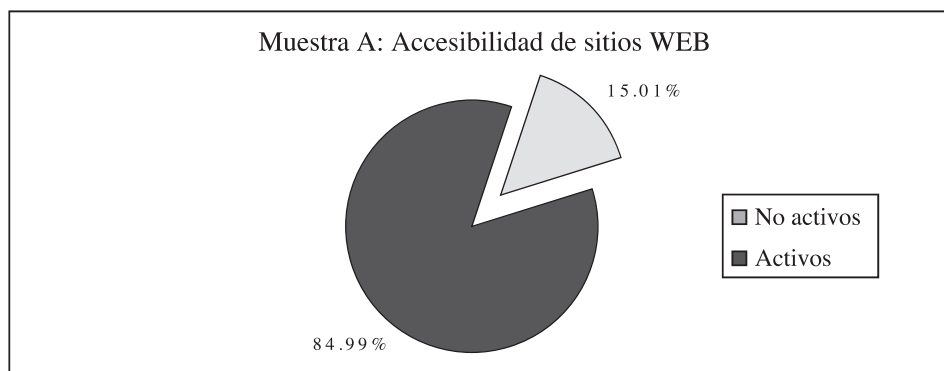
Ciclo de vida de los sitios WEB

En el campo de las publicaciones impresas los periodos vigencia de la información se encuentran establecidos en proporción al avance de la ciencia en los diferentes campos, después de ese período de tiempo la publicación adquiere otro tipo de valor: científico-cultural e histórico. Sin embargo en Internet este periodo de vida se encuentra limitado por la permanencia del mismo en el soporte de almacenamiento en forma física un servidor, sujeto al cambio de tecnología, fallas electrónicas de comunicación y otros. Su vida útil es determinada por elementos tecnológicos sin valorar su aporte científico-cultural.

La vida de un sitio Web puede llegar a ser efímera y la característica de ser un documento casi único, sin disponibilidad de copias pero de acceso universal, excede la concepción tradicional de un soporte documental y va más allá que el simple almacenamiento de un documento electrónico.

En este proceso de generación de datos métricos se cuantifico la inactividad de sitios Web. La muestra A., después de 15 meses mostró resultados de inactividad que acumulativamente representa el 15.01 por ciento de sitios Web. El 84.9 por ciento de los sitios existen físicamente pero no necesariamente es información que permanece intacta (ver gráfico # 6).

Gráfico # 6

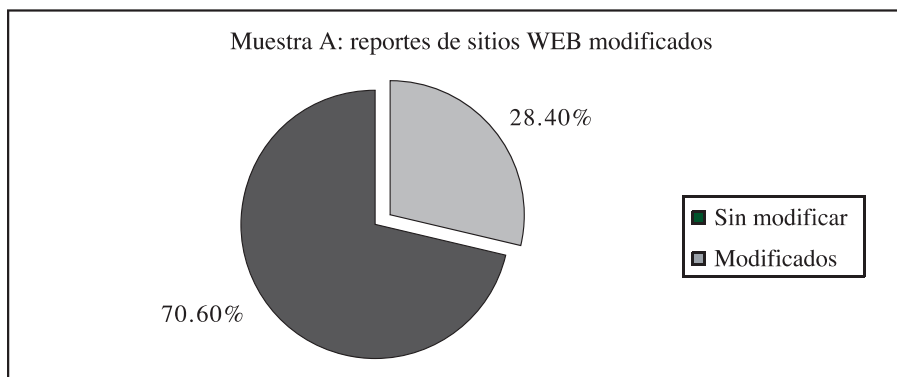


Fuente: Datos métricos proyectos de graduación 2003.

Se encontró en el grupo de sitios Web activos. El 28.4 por ciento de sitios activo hace referencia a información que fue modificada al menos una vez. Uno de sus detalles al ser modificado en

forma o contenido hace dudar sobre la veracidad del documento y el principio de originalidad de la fuente de información. (ver gráfico # 7).

Gráfico # 7

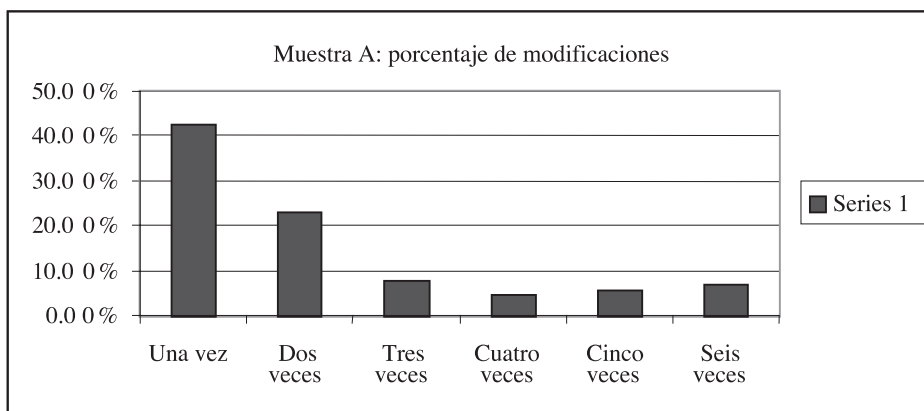


Fuente: Datos métricos proyectos de graduación 2003.

El 28.4 por ciento de la información ha sufrido modificaciones, no es igual a la consultada por los autores en el momento de ser citada e incorporada en sus trabajos. Sin embargo este grupo de sitios Web presento reiteradas modificaciones en el transcurso de los 15 meses de estudio. El 22.82 por ciento reporto haber sido modificado dos veces. El 7.60 por ciento reporto haber sido modificado tres veces. El 4.30 por ciento

reporto haber sido modificado cuatro veces. El 5.40 por ciento reporto haber sido modificado cinco veces y el 6.50 por ciento reporto haber sido modificado seis veces. En el grupo de sitios Web que reportaron ser modificado seis veces se encontró que correspondían a sitios interactivos que se forman con elementos de otras páginas y otros recursos en el momento de ingresar a ellos (ver gráfico # 8).

Gráfico # 8

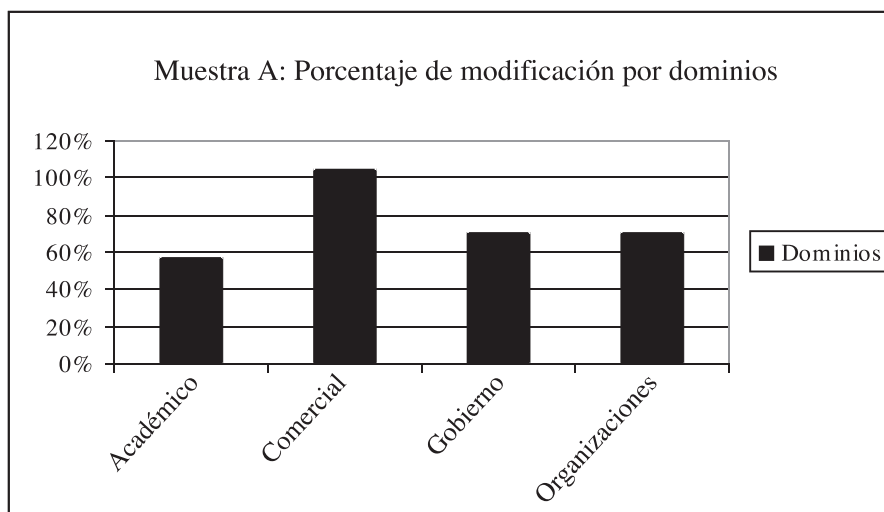


Fuente: Datos métricos proyectos de graduación 2003.

La variable de sitios modificados por dominios mostró el siguiente resultado. La categoría de sitios comerciales representa al grupo con mayores modificaciones recibidas con un 104 por ciento. Las categorías de dominios de

Gobierno y Organizaciones del segmento de sitios Web que presentaron modificaciones es un 70 por ciento y el la categoría de sitios académicos presento una menor modificación de la información con 57 por ciento (ver gráfico # 9).

Gráfico # 9



Fuente: Datos métricos proyectos de graduación 2003

Conclusiones

Los estudios sobre el comportamiento de los sitios Web existentes se enmarcan en mucho en lograr posicionarlos y mejorar su recuperación. Los factores de conectividad de la red, sus modificaciones, la conservación de información es poco cuestionada. Internet II como recursos para la investigación será la plataforma adecuada para la conservación del patrimonio científico-cultural, la red como la conocemos hoy no es el soporte adecuado para tal fin y acarrea implicaciones importantes para un proceso académico o de investigación si no se toman las previsiones correspondientes.

Las poblaciones estudiadas presentan una adecuada proporción sobre la utilización de fuentes de información con relación a su soporte. Los sitios Web pretenden dar actualización a los procesos de investigación y su citación es un buen recurso para lograr este objetivo.

La intervención de elementos tecnológicos como la variabilidad de conexión, el mantenimiento de páginas Web y la imposibilidad de retener información relevante en el tiempo en Internet, han sido hasta fecha variables no registradas. Este es el aporte de este trabajo, proporcionar datos de este fenómeno hasta ahora poco cuantificado.

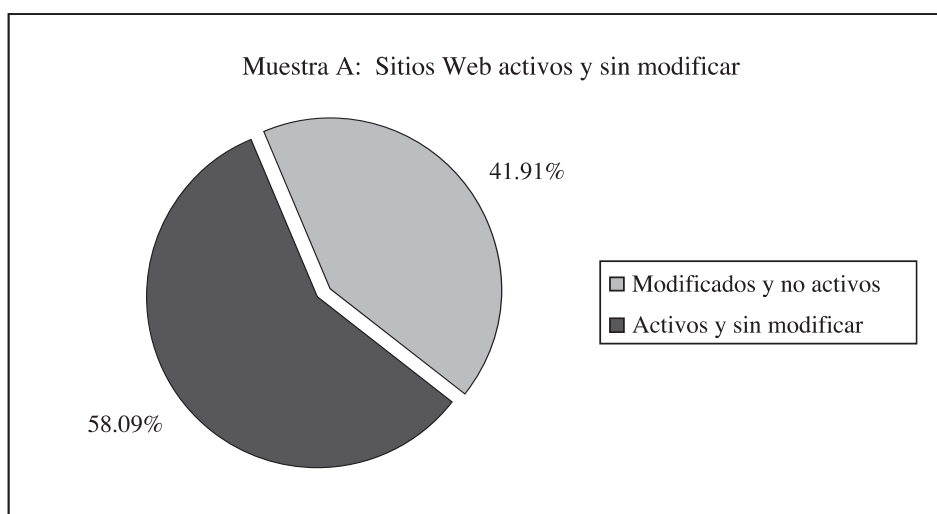
Los datos mostrados permiten construir una idea de la inactividad de sitios Web en un lapso de tiempo establecido. En la muestra A, representa un 15 por ciento del total de sitios que se entraron en inactividad y los sitios modificados su porcentaje es el 28.4 por ciento con relación al grupo de sitios Web que existen y permanecen físicamente.

Si se ponderan adecuadamente las variables de sitios inactivos y la variable sitios que se han modificado a la fecha, se puede afirmar que en la muestra A. El grupo de sitios Web que son iguales a los consultados por sus autores en el

proceso de elaboración de su proyecto de graduación después de 15 meses es del 58.09 por ciento. El grupo integrado por los sitios Web

que presentaron inactividad y modificación es del 41.91 por ciento porcentaje absoluto de la muestra total (ver gráfico 10).

Gráfico # 10



Fuente: Datos métricos proyectos de graduación 2003.

El grupo de sitios que se mantienen iguales a los consultados es mayor a los que mostraron inactividad y modificaciones. Pero evidencia que el segmento de inactivos y modificados es alto.

Los factores internos de funcionamiento de esta red, suman un factor de incertidumbre en la conservación de la información para sustentar un proceso de investigación. Recomendando a los autores realizar su propio respaldo de información o establecer un procedimiento que garantice su conservación.

El mecanismo para asegurar la calidad de citación de sitios Web pasa inevitablemente por la formación de parámetros para su validación entre ellos y los más destacados deben ser:

a. Realizar la citación de información de sitios Web de acuerdo a la normativa iso 690 que permite describir adecuadamente un recurso electrónico. La mayoría de estilos de citación bibliográfica utilizados en la investigación lo incorporan pero se debe verificar que ellos cumplan con la normativa.

- b. Establecer parámetros de validación para los dominios Web acordes con su función social, económica, política e ideológica, verificar que sea una fuente de información confiable y relevante para el campo de estudio.
- c. Instaurar una política de respaldo de información en diferentes formatos que también puede incluir los formatos tradicionales y el depositarlos en un sistema de información que garantice su conservación.
- d. Establecer una directriz de políticas de investigación que asegure la calidad de los recursos de investigación y exhaustividad en las fuentes de información para evitar la segmentación de recursos de investigación.
- e. Capacitar a los investigadores en la importancia de distinguir el momento oportuno y preciso de respaldar el desarrollo de su investigación con sólidas referencias bibliográficas que pueden cimentar su permanencia

en el tiempo, que garantice un aporte real al conocimiento y engrandecer el patrimonio científico-cultural.

- f. Impulsar en el ámbito académico y científico el desarrollo de una cultura de registro de toda la producción electrónica con la normativa internacional del control bibliográfico ISBN y ISSN.

Con los mecanismos de aseguramiento de la calidad de citación en procesos de investigación y requisitos académicos anteriormente expuestos se fortalecerá los procesos de adquisición y desarrollo de habilidades informativas. Impulsará el camino de la alfabetización informativa con bases sólidas y efectividad en el uso de los recursos de información.

Bibliografía consultada

- Abaitua, J. 1999. *Sobre Internet: generalidades*. Bilbao, Universidad de Deusto. Facultad de Filosofía y Letras. <http://sirio.deusto.es/abaitua/konzeptu/internet/htm>
- Araújo Ruiz, J. A. y J. Arencibia. "Infometría, bibliometría y cienciometría: aspectos teórico-prácticos". *Rev. ACIMED* 10 (4)154, 2002
- Bunge, M. 1985. *La ciencia: su método y filosofía*. Buenos Aires., Siglo XXI.
- Carrizo Sainero, G. "Hacia un concepto de Bibliometría". *Rev. Journal of spanish Research on Information Science* 200 1(2) 6583-6593. 2001.
- Dellavale, R. P., et al. 2003. "Going, Going, Gone: Lost Internet reference". *Rev. Science* 302 (5646), 787- 788 p.
- Gómez Sustaita, R. 2005. "La enseñanza de las habilidades informativas". En: *Revista Postgrado-Universidad Autónoma de Guadalajara*. <http://gnesis.uag.mx/postgrado/revistaelect/educa/eduoo1.htm>
- Goñi Camejo, I. 2000. "Algunas reflexiones sobre el concepto de información y sus implicaciones para el desarrollo de las ciencias de la información". *Rev. ACIMED*. 8(3), 201- 207 p.
- Hernández Sampieri, R. 2000. *Metodología de la investigación*. 2ed. México, D.F. McGraw-Hill, 501p.
- Meza Pérez, J. P. 2004. *Datosmétricos: Proyectos de graduación 2003 de la Universidad EARTH*. Guácimo. EARTH. 34 p.
- Polanco, X. "Infometría e ingeniería del conocimiento: exploración de datos y análisis de la información en vista del descubrimiento de conocimientos". In. Jaramillo, H y M. Albornoz (eds). *El Universo de la medicina*. Bogotá, TM Editores. 335-350 p.
- Urbizagástegui Alvarado, R. 1999. "La ley de Lotka y la literatura bibliométrica". *Rev. Investigación Bibliotecológica* 13(27) 125-141 p.