

Medidas de popularidad de la educación superior cubana: un análisis webométrico

Por Yeter Caraballo-Pérez, Verena Torres-Cárdenas, Aida-C. Noda-Amoros y Javier-A. Herrera-Toscano



Yeter Caraballo-Pérez es graduada en bibliotecología y ciencias de la información en la Universidad de La Habana. Trabaja en el Centro de información del Instituto de ciencia animal, donde investiga sobre bases de datos, bibliotecas digitales y virtuales y arquitectura de información, y ha impartido seminarios sobre arquitectura de información, metadatos y visibilidad de la información en la web.



Verena Torres-Cárdenas es licenciada en matemáticas por la Universidad de La Habana en 1970, doctora en ciencias agrícolas en 1985 y doctora de segundo nivel en 2004. Trabaja en el Instituto de ciencia animal como investigadora titular, donde ha realizado múltiples trabajos de investigación. Es profesora auxiliar adjunta y directora del Departamento de biomatemática.



Aida-Carmelita Noda-Amoros es técnico en análisis matemático y trabaja en el Departamento de biomatemática del Instituto de ciencia animal desde 1970. En su trayectoria laboral ha contribuido a la realización de múltiples trabajos científicos relacionados con las ciencias agropecuarias, en los cuales ha asesorado y orientado a los investigadores.



Javier-A. Herrera-Toscano es ingeniero agrónomo desde 2001, investigador agregado del Departamento de manejo y alimentación de rumiantes, perteneciente al Instituto de ciencia animal, y máster en ciencias en 2005. Ha publicado artículos y ha impartido cursos sobre ovino-caprino y alternativas de alimentación para el trópico. Su principal línea de investigación se enfoca a la optimización de sistemas agropecuarios.

Resumen: Se presentan algunas consideraciones sobre la popularidad de los centros del Ministerio de Educación Superior (MES) de Cuba. Se esboza el papel que desempeña la visibilidad en la world wide web y su vínculo e importancia en el posicionamiento, pues los motores de búsqueda utilizan el análisis de enlaces (o análisis hipertextual) para tratar de inferir no solamente los temas, sino la calidad o la posible condición de “autoridad” de un sitio web. Tomando esto como base se realiza un estudio webométrico de los 17 centros que tienen sitios web y se determina la relación entre el número de enlaces y el tráfico de las páginas de los centros del MES a partir del análisis de los buscadores Google, AlltheWeb y Altavista, de Internet archive y el ranking de Alexa. Se obtienen cuatro resultados: 1) los centros de MES tuvieron un mejor comportamiento en el buscador Altavista ($EE=\pm 0,00210$) y de éstos el valor significativo fue el del centro 13 marcado con el superíndice a; 2) en la medida que se incrementaron las páginas archivadas en Internet archive al paso de los años, los sitios web de los centros del MES recibieron mayores citas o enlaces de los buscadores, en especial de Google; 3) los enlaces y el ranking de los centros del MES se comportaron significativamente mejor en el buscador Alltheweb para un EE de $\pm 0,23$ y un R^2 igual a 0,44%; 4) existen diferencias importantes en las páginas de los centros del MES en Internet archive respecto a los enlaces en el ranking de Alexa.

Palabras clave: Webmetría, Visibilidad, Tráfico, Sitio web, Centros del MES.

Title: Popularity measurements of Cuban higher education: a webmetric analysis

Abstract: We discuss the popularity of Cuba's MES (Ministry of Higher Education) centres, as well as their interconnectedness and visibility on the world wide web. The role of links in the positioning and visibility of these sites was studied because web browsers use link analysis (or hypertext analysis) to infer not only the content but the quality or the possible “authority” status of a web site. A webmetric study of the 17 centres with web sites was conducted. The relationship between the number of links and page traffic of the MES centres was determined by analysis of the browsers Google, Alltheweb, and Altavista, and also from the Internet Archive and Traffic Rank (Alexa ranking). Four results were obtained: i) the MES centres had better visibility in the browser Altavista, with a Standard Error (SE) $\pm 0,00210$; of these, a significantly better value was observed for the 13th centre (superscript a); ii) as page storage in the Internet archive increased, more links or cites were received from the browsers; iii) the links and ranking in Alltheweb showed significantly better performance ($SE \pm 0,23$) and a coefficient of determination (R^2) of 0,44%; iv) significant differences exist among MES centre sites in the Internet archive, with respect to links and activity reported in the Traffic Rank.

Keywords: Webmetric, Webometrics, Visibility, Traffic, Web site, MES centres.

Caraballo-Pérez, Yeter; Torres-Cárdenas, Verena; Noda-Amoros, Aida C.; Herrera-Toscano, Javier A. “Medidas de popularidad de la educación superior cubana: un análisis webmétrico”. En: *El profesional de la información*, 2008, julio-agosto, v. 17, n. 4, pp. 443-448.

DOI: 10.3145/epi.2008.jul.12

Introducción

LOS ASPECTOS RELACIONADOS CON LA POPULARIDAD de una web influyen enormemente en el aumento de su visibilidad en internet, de tal manera que, mientras mayor sea el número de enlaces que se reciben de otros sitios, mayor será su visibilidad, a lo que habría que sumar la calidad o pagerank que tengan los sitios que citan.

En la página de Google “*Tecnología Google*” se define visibilidad como la medida que indica el número de enlaces que un sitio web recibe de otros y además la calidad de los mismos, esto es, si esos sitios tienen a su vez una alta popularidad hipertextual (si son muy enlazados). Por lo tanto se trata de un rasgo que en parte es hereditario: un sitio web citado por otro muy popular tendrá mejor posicionamiento que uno citado por uno poco popular.

Los enlaces juegan un papel muy importante en el posicionamiento, ya que los motores de búsqueda los analizan (análisis hipertextual) para inferir no solamente los temas, sino la calidad o la posible condición de “autoridad” de un sitio web (Codina, 2005).

Álvarez (2006) también plantea que “existe una nueva generación de buscadores que ponderan los resultados de sus búsquedas en función de lo popular que sea un sitio. Éstos realizan un sencillo cálculo que consiste en contar los enlaces que tienen otras páginas hacia un sitio web determinado”. Aquellos que tengan más enlaces dirigidos hacia ellos tendrán una puntuación más alta y aparecerán más arriba en los resultados. Este es el algoritmo que utilizan buscadores como Google,

Alltheweb y Altavista para medir la popularidad de un determinado sitio y en los que esta contribución se centró para obtener el número de enlaces.

Por otra parte Alexa es una empresa que realiza análisis y mediciones sobre tráfico en internet y que proporciona también una base de datos y un directorio de webs. Usa una de las medidas de popularidad más conocidas, el *Traffic rank*, un cálculo que combina tanto el número de usuarios que visitan un dominio web como la cantidad de páginas vistas (Alexa, 2007).

Por su parte *Internet archive* es una organización que mantiene una colección de páginas web y materiales digitales y que lanzó la *Wayback machine*, por medio de la cual se puede constatar el comportamiento de la accesibilidad del sitio web en el pasado. En 1996 *Internet archive* comenzó la recopilación de estos documentos para su preservación, y en 2001 empezó a proporcionar el acceso a los 10.000 millones de páginas que tiene almacenadas (*Internet archive*, 2007).

Este estudio sobre el nivel de visibilidad de las webs de los centros del *Ministerio de Educación Superior (MES)* de Cuba, se basa en los anteriores recursos descritos como fuentes de información fiables. Según Codina (2005) la mayoría de los usuarios de motores de búsqueda sólo consultan los diez primeros resultados recuperados. Se determinará la relación entre el número de enlaces y el tráfico de las páginas de los centros del *MES* a partir del análisis de los buscadores *Google*, *AlltheWeb* y *Altavista*, de *Internet archive* y el ranking de *Alexa*, utilizando los criterios de enlaces.

Materiales y métodos

Para desarrollar este trabajo se identificaron todas las páginas de los centros que pertenecen al *MES*, se obtuvo el número de enlaces a partir de los buscadores (*Google*, *AlltheWeb* y *Altavista*) y el tráfico del ranking de *Alexa* de 2006 a 2007, y se analizó en *Internet archive* su comportamiento desde enero de 1996 hasta febrero de 2007. El *MES* comprende un total de 22 centros, de los cuales 17 tenían disponibles sus sitios web.

Se realizó un análisis estadístico para medir la relación entre las páginas de los centros del *MES* y el comportamiento del número de enlaces en cada uno de los buscadores, del tráfico en el ranking de *Alexa* y su disponibilidad en *Internet archive*.

Los datos se procesaron mediante el paquete estadístico *Infostat*. Se realizó un análisis de varianza según modelo de clasificación simple para conocer el efecto de los buscadores (*AlltheWeb*, *Altavista* y *Google*), *Internet archive*, y la interacción entre éstos en cada una de las páginas de los centros pertenecientes al *MES*.

Se obtuvo la frecuencia de las citas de los centros en los buscadores, con lo cual se condujo un análisis de proporciones, a través de las tablas de contingencia usando la dócima X^2 (*chi-cuadrado*). Se utilizó la dócima de comparación múltiple de **Duncan** (1955) para $p < 0,05$ en los casos en que se alcanzaron diferencias significativas.

Se usó el análisis de regresión no lineal para establecer la relación entre los indicadores estudiados, y por las características del modelo

usado se utilizó la transformación según logaritmo neperiano (\ln). Para el análisis se aplicó la siguiente fórmula: donde:

– y es el valor del buscador (*Alltheweb*, *Altavista*, *Google*).

– V_0 es la constante del modelo ajustado.

– x es el valor de la transformación según \ln del ranking de *Alexa*.

– b es el valor de acuerdo al modelo ajustado.

En este análisis se eliminaron los datos correspondientes a la frecuencia 0.

Resultados y discusión

Las observaciones en los buscadores muestran el número de enla-

ces o citas (tabla 1) que obtuvieron los centros que pertenecen al *MES*. Se observa que el menor error estándar (*EE*) es el del buscador *Altavista* con un valor de $\pm 0,00210$.

El número de enlaces de los centros del *MES* según *Altavista* oscilaba de 3 a 3,080. Según *Google* estaba entre 0 y 283.

En la tabla 1 se observa que en *Altavista* el centro que presentó el mejor valor fue el 13, marcado con el superíndice “a”, y le siguen consecutivamente los centros 6, 17 y 16. El resultado coincide con el nivel de reconocimiento que ha alcanzado el centro que ocupa el primer lugar tanto a nivel nacional como internacional y así se evidencia en el resto de los buscadores analizados.

Por su parte los correspondientes a los números 1, 9 y 15 (^{hi}); 4 y 7(^{ij}); 3 y 5(^j) tienen un comportamiento similar, aunque es válido destacar que están entre la media de los mejores más citados en *Altavista*.

En la figura 1 se presenta el efecto que tienen las páginas archivadas en *Internet archive* desde el año 1996 hasta el 2007 en el número de enlaces obtenido de los buscadores *AlltheWeb*, *Altavista* y *Google* de 2006 a 2007.

Los resultados de *AlltheWeb* y *Altavista* mostraron igual significación ($p < 0,01$), mientras que *Google* se comportó mejor para $p < 0,001$ y con menor *EE* de $\pm 1,06$. De este modo se observó una relación directamente proporcional, es decir, que en la medida en que se incrementaron las páginas archivadas en *Internet archive* al paso de los años, mayor cantidad de citas o enlaces recibieron los sitios web de los centros del *MES* de los buscadores, en especial de *Google*.

Se puede observar además que el valor del coeficiente de determinación que se obtuvo fue muy bajo en los tres análisis realizados, al parecer por los aspectos relacionados indirectamente con los indicadores estudiados, por ejemplo que la cobertura de la web visible es actualmente incompleta (**Aguillo et al., 2006**).

Al poner en *Internet archive* el comportamiento de las páginas de los centros del *MES* en función de las observaciones de los enlaces en *AlltheWeb*, *Altavista* y *Google*, se obtuvieron las siguientes ecuaciones lineales para estimar el efecto entre las páginas archivadas en *Internet archive* y el número de enlaces obtenido de cada uno de los buscadores:

– *AlltheWeb* = $76,18 + 17,58 * \text{Internet archive}$. $R^2 = 0,20$; $EE(b) = +0,20$.

– *Altavista* = $179,67 + 25,75 *$

Centros	<i>AlltheWeb</i> %		<i>Altavista</i> %		<i>Google</i> %	
1	0,00013	i	0,01654	hi	0,02532	gh
2	0,00026	g	0,02819	g	0,02240	ghi
3	0,00039	jk	0,00972	j	0,02434	gh
4	0,00052	ij	0,01397	ij	0,02824	gh
5	0,00065	lm	0,00819	j	0,01461	ghi
6	0,00078	f	0,14213	b	0,06037	e
7	0,00091	kl	0,01309	ij	0,00000	i
8	0,00104	m	0,00024	k	0,00779	hi
9	0,00117	hi	0,01726	hi	0,03505	fg
10	0,00130	hi	0,03638	f	0,05063	ef
11	0,00143	b	0,10198	d	0,11490	c
12	0,00156	ij	0,02208	hi	0,00000	i
13	0,00169	a	0,24733	a	0,27556	a
14	0,00182	e	0,04955	e	0,06816	de
15	0,00195	kl	0,02674	hi	0,01168	ghi
16	0,00208	c	0,11965	c	0,17332	b
17	0,00220	d	0,14695	b	0,08763	d
EE±	2,67967*		0,00210*		0,00734*	

abcdefghijklm: Letras distintas indican diferencias notables ($p \leq 0,05$) (**Duncan, 1955**)
*: $p < 0,001$

Tabla 1. Dinámica del comportamiento de los enlaces en los buscadores analizados desde 2006 a 2007

Internet archive. $R^2 = 0,18$; $EE(b) = +13,97$.

– Google = $-6,81 + 3,13 * Internet\ archive$. $R^2 = 0,37$; $EE(b) = +1,06$.

Los resultados obtenidos no difieren de los planteamientos expuestos en la literatura especializada, donde se pudo apreciar que la accesibilidad de un sitio web juega un papel primordial en su visibilidad, en tanto es la base para que el sitio sea accesible por la audiencia y pueda ser citado por otros.

Se estimó el efecto del ranking de Alexa en los enlaces obtenidos de los buscadores de 2006–2007 (figura 2), a partir de la obtención de las ecuaciones lineales que se muestran a continuación:

– AlltheWeb = $10,23 - 0,67 * ranking\ Alexa$. $R^2 = 0,44$. $EE(b) = +0,23$.

– Altavista = $10,36 - 0,61 * ranking\ Alexa$. $R^2 = 0,42$. $EE(b) = +0,22$.

– Google = $7,04 - 0,47 * ranking\ Alexa$. $R^2 = 0,35$. $EE(b) = +0,2$.

Los resultados mostrados en la figura 2 evidencian que los enlaces y el ranking de los centros del MES se comportaron significativamente mejor en el buscador AlltheWeb para un EE de $\pm 0,23$ y un R^2 igual a 0,44%.

Se observó una relación inversamente proporcional, pues a medida que aumentaba el ranking de los sitios, los buscadores AlltheWeb, Altavista y Google disminuían su número de citas de 7,09 a 4,34; de 7,5 a 4,99 y de 4,84 a 2,9 respectivamente.

Esta observación puede deberse al algoritmo utilizado en el ranking de Alexa para medir la popularidad de un sitio web, el cual está basado en la combinación del número de usuarios que visitan un mismo dominio web y del número de páginas vistas por los usuarios en ese domi-

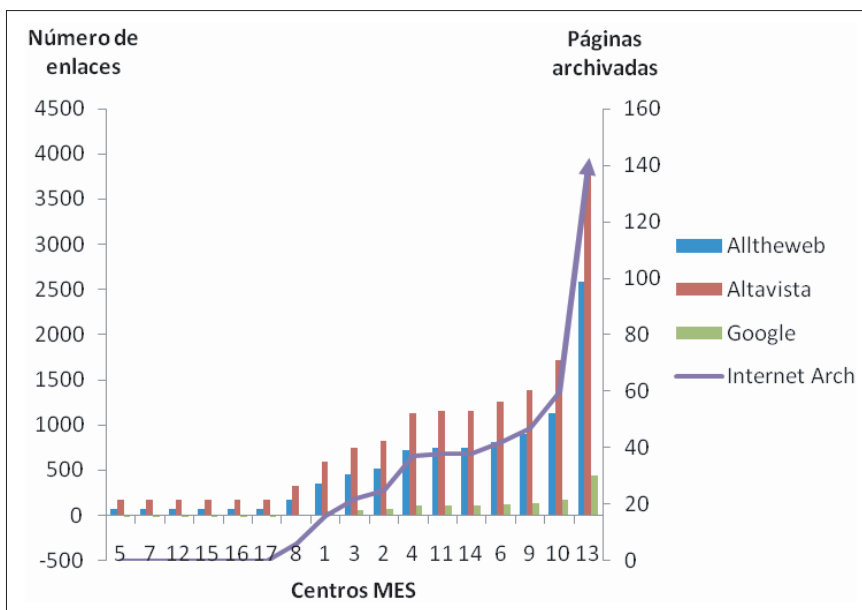


Figura 1. Efecto de las páginas archivadas en Internet archive (1996-2007) en los enlaces obtenidos de los buscadores (2006-2007). AlltheWeb y Altavista: $p < 0,01$ – Google: $p < 0,001$

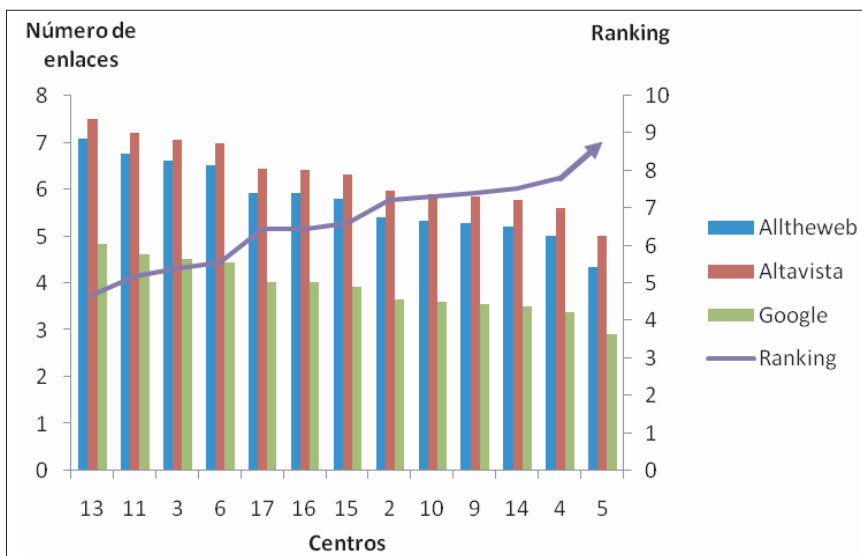


Figura 2. Efecto del ranking de Alexa en los enlaces obtenidos de los buscadores de 2006-2007 AlltheWeb, Altavista y Google: $p < 0,01$

nio. Este es un cálculo totalmente diferente al realizado por los buscadores, los cuales, en el caso que nos ocupa, cuentan los enlaces que tienen otras páginas hacia un sitio web determinado.

En los resultados expuestos en la tabla 2 se aprecia que existen diferencias notables en el número de páginas archivadas en Internet archive respecto al ranking de Alexa.

Esto se puede deber, por una

	Medias de páginas
Internet archive	27,88 ^a
Alexa	1.452,81 ^b
$EE (\pm)$	330,25**

^{ab} Letras distintas indican diferencias significativas ($p < 0,05$) (Duncan, 1955)
^{**} $p < 0,01$

Tabla 2. Relación entre las páginas archivadas en Internet archive (1996–2007) y el ranking de Alexa (2006–2007) de los centros del MES

parte, a la medida de *Wayback machine* reportada en el sitio web de *Internet archive* (2007). Esta medida -como se ha explicado- muestra los resultados del comportamiento de las páginas de determinado sitio web en el pasado. Así se sabe si una página ha estado inaccesible por determinadas razones; y por otra parte, por el algoritmo utilizado en el ranking de *Alexa*, el cual reporta la cantidad de citas que recibe un sitio web de otros (*Alexa*, 2007). Como se aprecia son dos criterios totalmente diferentes que influyeron en el resultado obtenido en la tabla 2, a pesar de lo cual es necesario aclarar que se deben tener en cuenta para medir la visibilidad de una web.

De los resultados se puede concluir que:

– Los centros de *MES* tuvieron un mejor comportamiento en *Altavista* ($EE = \pm 0,00210$), de los cuales, el mejor valor significativo fue el del centro 13 marcado con el superíndice a.

– En la medida que se incrementaron las páginas archivadas en *Internet archive* con el paso de los años, los sitios web recibieron mayores citas de los centros del *MES* de los buscadores, en especial de *Google*.

– Los enlaces y el ranking de los centros del *MES* se comportaron mejor en *AlltheWeb* para un EE de $\pm 0,23$ y un R^2 igual a 0,44%.

– Existen diferencias importantes en las páginas de los centros del *MES* en *Internet archive* respecto a los enlaces en el ranking de *Alexa*.

Consideraciones finales

Si se desea que un sitio alcance una buena popularidad y por ende una buena visibilidad, a partir del término que encierra todo el quehacer de la institución, por ejemplo “ciencia animal”, es necesario que tenga contenidos de calidad sobre

1	Censa: Centro nacional de sanidad agropecuaria
2	CNIC: Centro nacional de investigaciones científicas
3	ICA: Instituto de ciencia animal
4	INCA: Instituto nacional de ciencias agropecuarias
5	CUIJ: Centro universitario Isla de la juventud
6	Cujae: Instituto superior politécnico José Antonio Echeverría
7	CULT: Centro universitario de Las Tunas
8	CUSS: Centro universitario de Sancti Spiritus
9	UC: Universidad de Camagüey
10	UCF: Universidad de Cienfuegos
11	UCLV: Universidad central de Las Villas Marta Abreu
12	UDG: Universidad de Granma
13	UH: Universidad de La Habana
14	UMCC: Universidad de Matanzas Camilo Cienfuegos
15	UNAH: Universidad agraria de La Habana
16	UO: Universidad de Oriente
17	UPR: Universidad de Pinar del Río

Tabla de correspondencia de centros

este tema en todas sus páginas. Otra posibilidad es intentar que alcance buenas posiciones a partir de:

– estudios como la presente investigación y conocer el lugar donde se encuentra;

– gracias a la incorporación de metadatos que describan realmente la temática;

– a través de conseguir muchos enlaces, a partir de la colaboración con instituciones de prestigio sobre la misma temática.

Se deben tener en cuenta otros factores, como son las actualizaciones frecuentes del mismo, el mero resultado de una larga permanencia en la web, el mantenimiento del mismo dominio a lo largo del tiempo y una política de revisión y actualización frecuente de los enlaces, algo que en este estudio se pudo apreciar pues hubo sitios que no tuvieron una buena permanencia desde 1996 hasta la actualidad,

como es el caso de los centros 15, 16 y 17.

Bibliografía

Aguillo, Isidro F.; Ortega, José L.; Granadino, Begoña. “Contenidos del buscador Google. Distribución por países, dominios e idiomas”. En: *El profesional de la información*, 2006, septiembre-octubre, v. 15, n. 5, pp. 384–389.

Alexa. The web information company. 2007. Consultado en: 18-01-07. <http://www.alexa.com/>

Altavista. Acerca de Altavista. 2007. Consultado en: 20-01-07. <http://es.altavista.com/about/>

Álvarez, Miguel-Ángel. *Popularidad de un sitio web.* 2006. Consultado en: 25-12-07. <http://www.desarrolloweb.com>

Arroyo, N.; Ortega, J. L.; Pareja, V.; Prieto, J. A.; Aguillo, I. F. “Cibermetría. Estado de la cuestión”. En: *9^{as} Jornadas españolas de documentación, Fesabid* 2005.

Codina, Lluís. “Posicionamiento web: conceptos y ciclo de vida”. En: *Anuario Hipertext.net*, 2004, mayo. Consultado en: 18-01-07. <http://www.hipertext.net>

Codina, Lluís; Marcos, Mari-Carmen. “Posicionamiento web: conceptos y herramientas”. En: *El profesional de la información*, 2005, marzo-abril, v. 14, n. 2, pp. 84-99.

Duncan, David B. "Multiple range and multiple F tests". En: *Biometrics*, 1955, v. 11, n. 1.

Google. *Tecnología de Google.* 2007. Consultado en: 18-01-07.
<http://www.google.com/corporate/tech.html>

Infostat. *Software estadístico. Manual de usuario, Versión 1.* Córdoba, Argentina, 2001.

Internet archive. *Wayback machine.* 2007. Consultado en: 18-01-07.

<http://www.archive.org/>

Moreno-Pelayo, Valentín. "Interacción entre medidas de popularidad en el posicionamiento web". En: *El profesional de la información*, 2005, marzo-abril, v. 14, n. 2, pp. 100-107.

Yeter Caraballo-Pérez, Verena Torres-Cárdenas, Aida-C. Noda-Amoros, Javier-A. Herrera-Tosca-

no, Instituto de Ciencia Animal, Carretera Central Km. 47 1/2. Apartado Postal 24 CP 32700, San José de las Lajas, La Habana, Cuba.

jeter@ica.co.cu
vtorres@ica.co.cu
anoda@ica.co.cu
jherrer@ica.co.cu



ANUARIO

Think

EPI

Artículos en la sección GESTIÓN DE INFORMACIÓN

Gestión de contenidos 2.0

Por Jesús Tramullas

Documentos electrónicos: Reflexiones sobre las relaciones con la tecnología de los profesionales de la información. Parte I

Por Carlota Bustelo Ruesta

Documentos electrónicos: Reflexiones sobre las relaciones con la tecnología de los profesionales de la información. Parte II

Por Carlota Bustelo-Ruesta

Calidad en el servicio de las bibliotecas públicas: ¿Hablamos todos de lo mismo?

Por Roser Lozano

Hemeroteca digital en la biblioteca pública

Por Javier Guallar y Ernest Abadal

Gestión de información, crisis y emergencias

Por Ricardo Eito-Brun

Más información en <http://www.thinkepi.net/repositorio>

Ya puedes pasarnos tu pedido: epi@sarenet.es

Anuario ThinkEPI 2008

89,42 € + IVA = 93 €

Anuario ThinkEPI 2007 + 2008

115,39 € + IVA = 120 €

Anuario ThinkEPI

<http://www.thinkepi.net>