

La salud de las Web universitarias españolas

Mike Thelwall

*School of Computing and Information Technology, University of Wolverhampton,
Wulfruna Street, Wolverhampton WV1 1SB, UK. E-mail: m.thelwall@wlv.ac.uk*
Tel: +44 1902 321470 Fax: +44 1902 321478

Isidro F. Aguillo*

Laboratorio de Internet. CINDOC-CSIC. Joaquín Costa, 22. 28002 Madrid. España.
isidro@cindoc.csic.es Tel.: +34 91 5635482 Fax: +34 91 5642644

(*autor al que dirigir la correspondencia)

Resumen

La web se ha convertido en una importante herramienta para las universidades, donde se utiliza en una amplia variedad de formas, tales como publicar y diseminar actividades y resultados de investigación, proporcionar información administrativa y académica de interés para los estudiantes o facilitar la colaboración con otras instituciones nacionales e internacionales. Pero, ¿cómo están realmente utilizando la Web las universidades españolas y qué información se puede obtener sobre sus modos de comunicación en línea a través del estudio de los enlaces web?. Para obtener respuestas se han investigado 64 sedes Web de universidades utilizando un robot especializado, analizando posteriormente los datos obtenidos. Existe una gran variedad en el tamaño de las sedes universitarias, que reciben enlaces externos de otras sedes en proporción aproximada al tamaño de sus sedes. Aunque la Web académica española se situó por detrás de los cuatro países con las que se comparó, la mayoría de los enlaces que recibe provienen de dominios nacionales e internacionales de países no hispanohablantes lo que indica un amplio impacto internacional y un elevado grado de multilingüismo por parte de los autores de las web. Las páginas más enlazadas son mayormente aquellas que atraen enlaces generados automáticamente, con la sorprendente inclusión de algunos organismos gubernamentales.

Palabras clave: universidades españolas, robot, presencia en el web, sedes web, análisis cuantitativo, impacto internacional

A Health Check for University Use of the Web in Spain

Abstract

The Web has become an important tool for universities, and one that is employed in a variety of ways. Examples are: disseminating and publicising research findings and activities; publishing teaching and administrative information for students; and collaborating with other institutions nationally and internationally. But how effectively are Spanish universities using the Web and what information can be gained about online communication patterns through the study of Web links? This paper reports on an investigation into 64 university Web sites that were indexed using a specialist information science Web crawler and analysed using associated software. There were a wide variety of sizes for university Web sites and that universities attracted links from others broadly in proportion to their site size. The Spanish academic Web was found to lag behind those of the four countries that it was compared to. However, the most commonly targeted top-level Internet domains were from non-Spanish speaking high

Web using countries around the world, showing a broad international perspective and high degree of multilingualism for Web authors. The most highly targeted pages were mainly those that attracted automatically generated links, but several government ministries were a surprise inclusion.

Keywords: Spanish universities, crawler, web sites, web presence, quantitative analysis, international impact

Introducción

El continuo crecimiento de la web refleja su creciente presencia en cada vez mayor número de aspectos de la sociedad. Desde su pionero uso en la educación, ha encontrado firme partidarios en el gobierno y en la industria (1) así como su uso para propósitos sociales o recreativos. Pero sigue siendo el sector académico su principal usuario, como evidencia la importante proporción de páginas web en todo el mundo. Dentro de las universidades, la web es usada por una amplia variedad de razones (2) reflejando a la vez propósitos académicos y no académicos (3; 4). Dado el papel tan importante que juega la colaboración internacional en la investigación, y teniendo en cuenta que la web facilita dicha colaboración (5), es vital para la salud de las investigaciones de cualquier país, que sus académicos sean capaces de hacer un uso efectivo de la Web. Está claro, sin embargo, que en algunos países la Web esta comparativamente infrautilizada, y una revisión del uso y del perfil de la Web académica nacional debería ser un paso lógico para cualquier gobierno.

En España el esfuerzo de evaluación de la actividad universitaria (ANECA: www.aneca.es; Consejo de Coordinación Universitaria: <http://www.mec.es/consejou/calidad>) no ha proporcionado aún valores cuantitativos referentes al uso de Internet y la presencia en el Web en las distintas Universidades (6) a pesar del énfasis en las Nuevas Tecnologías que proponía el informe Universidad 2000 (7).

Sin embargo, parece lógico asumir de las estadísticas de uso de Internet en España (8) que las cifras, aunque muestran una evolución positiva, están aún lejos de acercarse a los parámetros que caracterizan a otros países europeos.

Otro factor importante a tener en cuenta es la oportunidad que la web proporciona para extraer información sobre los modelos de comunicación académica informal, mediante la investigación de los destinatarios de los enlaces Web. ¿Cuán internacional es la perspectiva de los autores de las páginas web académicas?, ¿A qué tipo de recursos se dirigen? ¿Qué países tienen las conexiones hipertextuales más próximas a España?

Antecedentes

Existen muchos estudios previos, donde se ha discutido sobre diferentes aspectos del uso de la Web académica, aunque ninguno ha tenido el objetivo explícito de realizar una revisión de un solo país. El tamaño de la web es probablemente un indicador demasiado básico para la mayor parte de los propósitos académicos, pero los enlaces Web atraen la atención, en parte por las analogías con las citas bibliográficas de las revistas científicas.

Anteriores trabajos, procedentes de un grupo especializado de científicos de la información, han creado los cimientos para el análisis cuantitativo de la Web. Los primeros documentos establecieron su potencial y el tipo de análisis que sería posible (9, 10, 11, 12, 13, 14, 15). Las posibilidades de estas aproximaciones han sido subrayadas posteriormente en los tratamientos teóricos de otros autores (16, 17, 18).

El principal tema de la investigación inicial de enlaces Web académicos fue si el recuento de enlaces tenía alguna validez para evaluar el impacto de las investigaciones o establecer modelos del uso de la Web académica (19, 20, 21, 22, 23, 24). Aunque esto no está totalmente resuelto, se ha descubierto que el recuento de enlaces a páginas Web de universidades o departamentos está significativamente asociado con los valores de la productividad investigadora en determinados países y disciplinas (23, 25, 26, 27, 28). Casi paradójicamente, investigaciones paralelas han mostrado que, por lo menos en el Reino Unido, menos de 1% de enlaces entre sedes Web de universidades lo eran entre publicaciones académicas formales, tales como versiones electrónicas de artículos de revistas (4). Las páginas más frecuentemente enlazadas son normalmente las páginas principales de las universidades (29), que no ofrecen casi ningún contenido semántico. A pesar de esto, casi el 90% de los enlaces que se encontraron, se crearon por razones relacionadas de alguna forma con las actividades académicas (4). Un enlace típico podría crearse desde una página de un grupo de investigación a la página principal de un grupo similar o colaborador, o al departamento, o la universidad que les da sede. Dicho esto, hay una gran variedad de razones para la creación de enlaces. Un recuento de enlaces a/o desde una institución se puede describir como la suma de una serie de prácticas de comunicación académica a través de la red de carácter informal con ciertas trazas de actividades de comunicación offline.

Además de la investigación específica dentro de dominios académicos, se han realizado investigaciones relacionadas con las herramientas usadas para la recopilación de datos. Investigaciones anteriores usaban interfaces de interrogación avanzadas de motores de búsquedas comerciales para obtener recuentos de páginas y enlaces. Investigaciones posteriores señalaron que estos generaban resultados altamente impredecibles e inestables (30, 31, 32), pero su desarrollo posterior parece haber sido mucho más estable. (33, 34). No obstante, existe un problema fundamental en su uso: los algoritmos de los robots se mantienen en secreto y así la fuente es poco fiable. Sin embargo, es necesario su uso para algunos tipos de datos dado que recorrer toda la Web cada vez que se necesita responder una pregunta concreta parece claramente impracticable.

Una respuesta al problema de los motores de búsqueda ha sido el uso de robots creados por o bajo el control de los investigadores (35, 36). Esto es válido para dominios suficientemente pequeños donde la recopilación sea realmente viable. Además proporciona más control sobre la obtención de los datos, permitiendo, por ejemplo, utilizar métodos de recuento alternativos (26, 37). Cuando se contabilizan los enlaces a/o desde un site Web es necesario agrupar los enlaces repetidos. Por ejemplo, si el propósito del recuento de enlaces es la evaluación de la diversidad de las direcciones objeto de enlace por un autor Web, los resultados pueden carecer de sentido si dicho autor ha creado una colección de cientos de páginas con enlaces idénticos. Una forma de evitar esto es ignorar el recuento de más de un enlace entre los mismos pares de dominios o de directorios. Esto se conoce como modelos de documentos alternativos de dominios y directorios (ADMs) respectivamente y se ha probado con éxito en el Reino

Unido (26). Bajo este sistema se agrupan todas las direcciones que tienen en común un dominio o un directorio para el posterior análisis.

El problema irresoluble y permanente es que un robot solo puede encontrar páginas mediante el seguimiento de enlaces o entradas humanas (38) y una parte de las páginas Web no se podrán localizar (“páginas huérfanas”). El resto de las páginas forma el conjunto de indexación pública (39), y los investigadores y los lectores deben aceptar que cualquier resultado publicado aplicará estos. En particular se omitirá la llamada Internet invisible (40) que consta de todos los contenidos que no pueden ser indexados correctamente por los motores de búsqueda, bien porque se accede a ellos a través de una pasarela, están en formato no textual o asimilable o requieren una palabra de paso o algún tipo de desprotección.

Cuestiones a resolver

El primer grupo de cuestiones lo constituyen la revisión de la salud de la Web académica española.

- ¿Es el tamaño de las sedes Web de las universidades españolas similar a aquellos de los países desarrollados?
- ¿Es el grado de interconexión entre las sedes Web de las universidades españolas aproximadamente similar al de otros países de su entorno?

El segundo grupo de preguntas se centra en temas relacionados con los modelos de comunicación académica informal, tal como se pueden identificar a partir de los enlaces Web.

- ¿Están los enlaces web repartidos uniformemente entre las universidades?, o ¿son una minoría los que atraen el mayor número de enlaces?
- ¿Existe una relación directa entre el número de enlaces a una universidad y el tamaño de su site Web?
- ¿Son los ADMs útiles para el análisis de enlaces en la Web académica española?
- ¿Qué países y que principales dominios Web atraen la mayoría de enlaces desde las universidades españolas?
- ¿Qué tipo de páginas son objeto preferente de enlace desde las universidades españolas?.

Metodología:

La lista de sedes Web de las Universidades se encuentra disponible tanto en directorios internacionales (<http://geowww.uibk.ac.at/univ/world.html>) como españoles (<http://www.rediris.es/recursos/centros/univ.es.html>). Un robot especializado de diseño propio (35) fue utilizado para recopilar la información de las sedes universitarias identificadas (Julio de 2002). En algunos casos los nombres de los dominios eran incorrectos o habían cambiado y las correcciones necesarias se realizaron manualmente. El robot cubrió en cada caso, todas las páginas que estuvieran bajo el mismo dominio institucional, en cualquiera de los dominios universitarios oficiales canónicos. Por ejemplo, el dominio de la Universidad de Salamanca (www.usal.es) incluía muchas direcciones diferentes aunque todas ellas acaban en .usal.es. Se vigiló durante el proceso

de recopilación que el programa no descargará sitios copiados (“mirrors”) o colecciones de páginas duplicadas de terceras. Además, se eliminaron los archivos de correo Web y las listas de distribución. La razón de todas estas acciones es que el principal objeto de este estudio son las páginas web creadas por la institución hospedadora. Por la naturaleza de la labor, con millones de páginas procesadas y un complicado método de seguridad para la identificación de áreas superfluas, solo las mayores colecciones de páginas se reconocerían durante el proceso de recopilación. Una vez finalizada la recolección, se analizó de nuevo las bases de datos de enlaces, usando programas de desarrollo propio que permiten identificar posibles colecciones de páginas superfluas que fueron vetadas y eliminadas manualmente. En esta segunda etapa se redujo un 20% el tamaño de la base de datos, lo que indica la magnitud del problema. Un ejemplo típico fueron las copias de grandes manuales de lenguajes de programación u otros softwares que tenían centenares de páginas.

Las bases de datos de enlaces producidas por el robot fueron procesadas por programas estadísticos para obtener un resumen de los resultados. Dichas bases de datos y el software utilizado para obtenerlas se encuentra disponible gratuitamente en la siguiente dirección: <http://cybermetrics.wlv.ac.uk/database>.

Resultados y discusión

Tamaños de las sedes

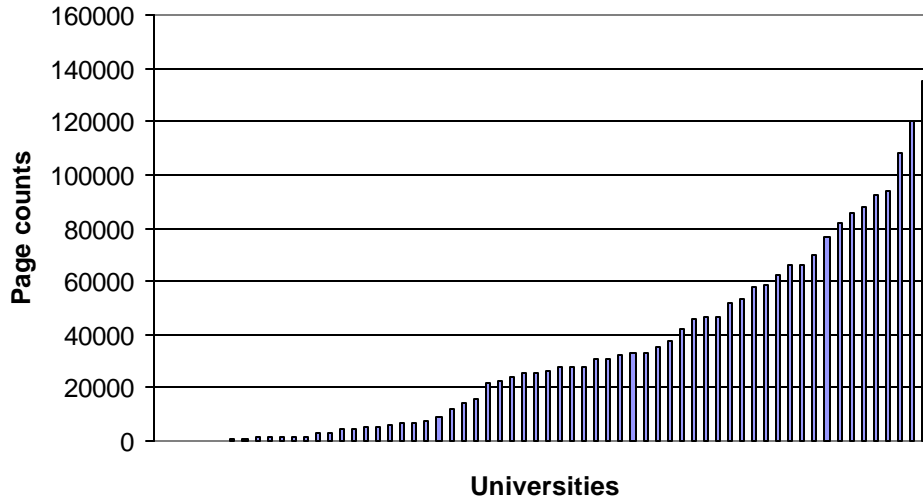
Como se indica en la Figura 1, el tamaño de las sedes varía enormemente, existiendo muchas de pequeño volumen. La mayor es la Universitat de Valencia. En la Tabla 2 se muestran las estadísticas básicas, junto con cifras similares extraídas de los países con los que se realiza la comparación, que es posible por que las bases de datos se obtuvieron usando el mismo robot (<http://cybermetrics.wlv.ac.uk/database>). Es difícil hacer comparaciones directas entre países, a consecuencia del tamaño y la variedad de los sistemas educativos. Por ejemplo en el Reino Unido hay un gran número de instituciones pequeñas además de los grandes centros. Los datos se han normalizado con el tamaño de población, pero la Web académica española parece estar ligeramente por detrás de otros países en términos de número de páginas per capita y apreciablemente en el caso de Australia. Parte de este retraso, sobre un 20%, podría ser atribuido al esfuerzo extra gastado en la identificación de páginas redundantes y superfluas para el conjunto de datos españoles. Pero se debe acabar indicando que el tamaño relativo de la Web es menor que el de los países con los que se ha comparado.

Tabla 1. Comparación de las estadísticas básicas de la Web de España en relación con países similares. Las figuras de población proceden del CIA World Factbook (<http://www.odci.gov/cia/publications/factbook>)

	España 2002	Australia 2001	Reino Unido 2002	Taiwan 2002	Nueva Zelanda 2002
Páginas universidades	2.123.943	2.726.475	5,438,941	2,150,690	315.142
Media de páginas por sede	25.917,5	64.153	29.757	20.737,5	38.081,5
Tamaño de la sede mayor	135.177	251.367	367.140	280.530	73.044
Universidades analizadas	64	38	111	46	8

Población	40.077.100	19.546.792	59.778.002	22.548.009	3.908.037
Tamaño de la Web per capita (España=100)	100	263	172	180	152

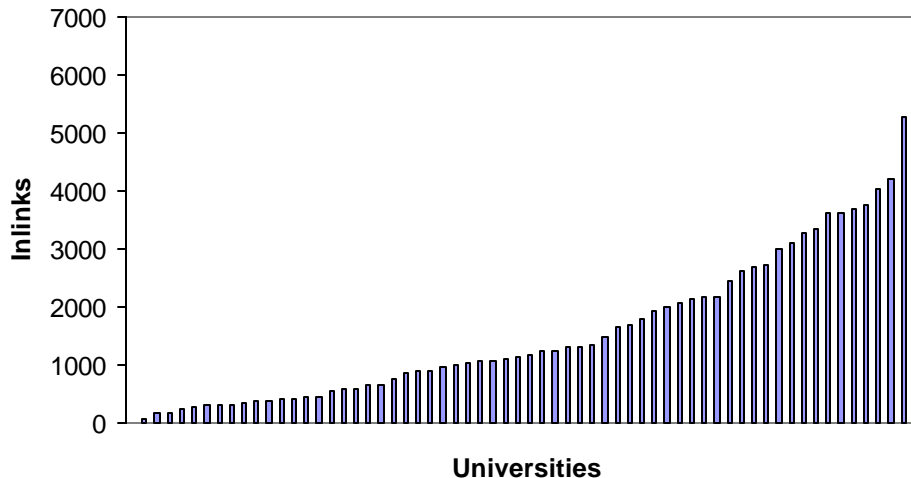
Figura 1. Gráfico donde se muestra la distribución de los tamaños (page counts) de las sedes Web de las 64 universidades españolas analizadas..



Enlaces entre universidades españolas

Según la Figura 2 el cómputo de enlaces recibidos es similar en las universidades españolas, aunque todas las universidades, menos una, atraen una cantidad razonable de enlaces. Si se tuviera que presentar un modelo similar al del Reino Unido (26), cabría esperar que el número de enlaces recibidos por cada universidad fuera aproximadamente proporcional a su productividad científica.

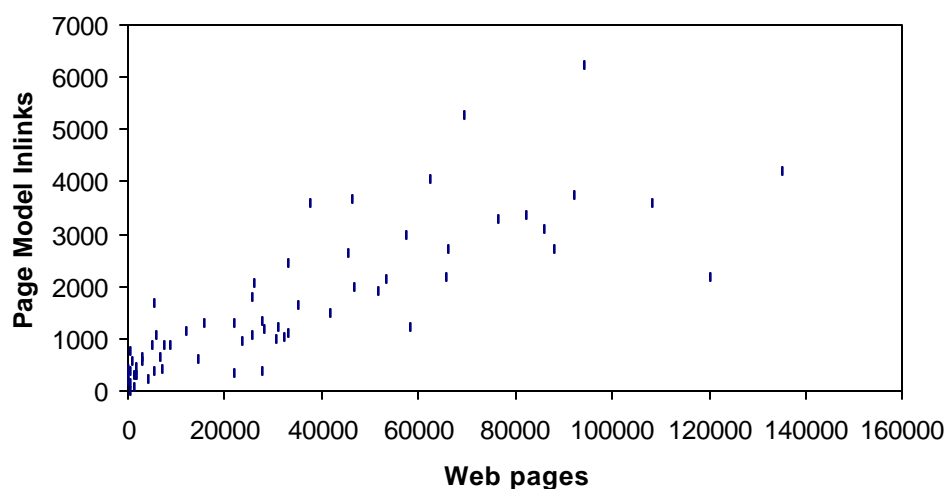
Figura 2. Gráfico donde se muestra la distribución del número de enlaces recibidos de sedes universitarias para las 64 universidades española recopiladas.



Relación entre el tamaño de las sedes y el recuento de enlaces recibidos.

La relación entre el tamaño de la sede y el recuento de enlaces recibidos se muestra en la Figura 3. Se puede observar que las grandes sedes tienden a atraer más enlaces. Si el modelo de interconexión está relacionado con la productividad investigadora, como en el Reino Unido entonces la desviación de una tendencia lineal puede representar instituciones relativamente productivas o improproductivas.

Figura 3. El gráfico muestra la relación entre el número de enlaces a una sede (“page model”) y su tamaño, medido en número de páginas.



El uso de ADMs

En el Reino Unido el recuento de enlaces entre dominios dió resultados más fiables (26) y puede comprobarse que los resultados del modelo dominio (Figura 5) están menos dispersos que los correspondientes al modelo fichero (Figura 3), a pesar de que en el primer modelo están implicados números menores lo que debería provocar una mayor dispersión (a igualdad del resto de los factores). Este resultado es un argumento a favor del modelo dominio.

La tendencia mayoritariamente lineal de las figuras 3 y 5 concuerda con el tamaño de los sedes Web, siendo este el principal factor responsable de la atracción de enlaces procedentes de otras universidades españolas. Los puntos aberrantes del gráfico son los casos donde se infringe esta regla. Según investigaciones previas, existen dos explicaciones posibles: 1) que la productividad científica de la institución sea mayor sea mayor o menor que la media nacional (26) ó 2) que las grandes colecciones puntuales de páginas puedan estar inflando artificialmente el tamaño de las sedes. Esta última explicación aparece reforzada en la comparación de las figuras 3 y 5, como ya se discutió anteriormente.

El dato fuera de lugar en el gráfico de directorios (Figura 4) es la Universidad de las Palmas de Gran Canaria, con 20,414 directorios y solo 1,019 enlaces recibidos a dichos directorios directorios. La causa de esto es la existencia de una enorme base de datos multi-directorio “Proyecto docente”. Otro punto sobre el eje y del gráfico de dominios es la Universidad de Gerona, que está aparentemente basada en un único dominio pero

que atrae 577 enlaces. Este es el resultado de un error en la recopilación: el dominio principal udg.edu fue recopilado, pero no así todos los subdominios acabados en udg.es, hecho que fue descubierto demasiado tarde para ser incluido en el robot, siendo factorizado dentro de los cálculos de los enlaces recibidos.

Figura 4. Gráfico donde se muestran las relaciones entre el número de enlaces desde un directorio único y el número de directorios únicos de la sede

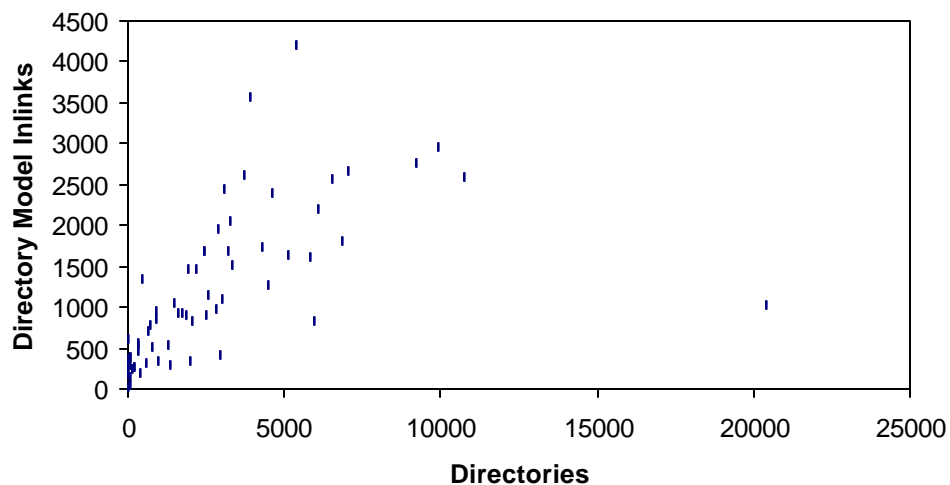
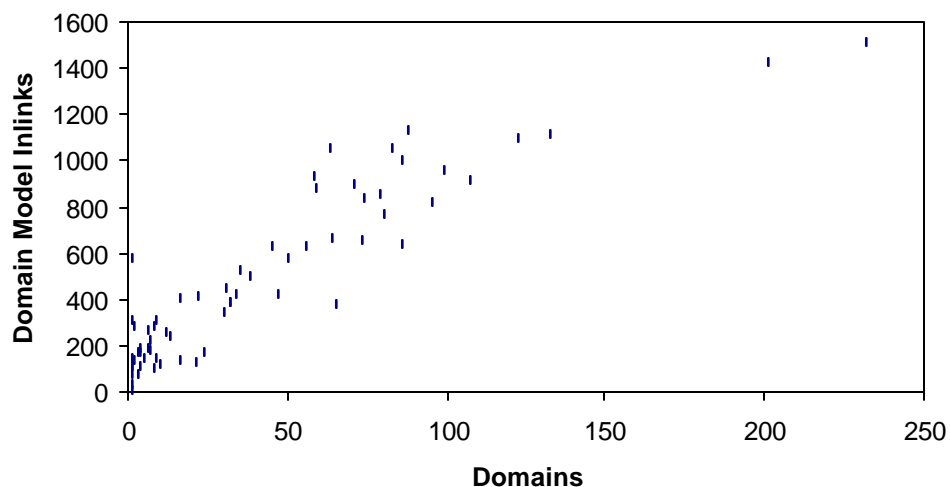


Figura 5. Gráfico donde se muestra las relaciones entre el número de enlaces desde un dominio único y el número de dominios únicos.

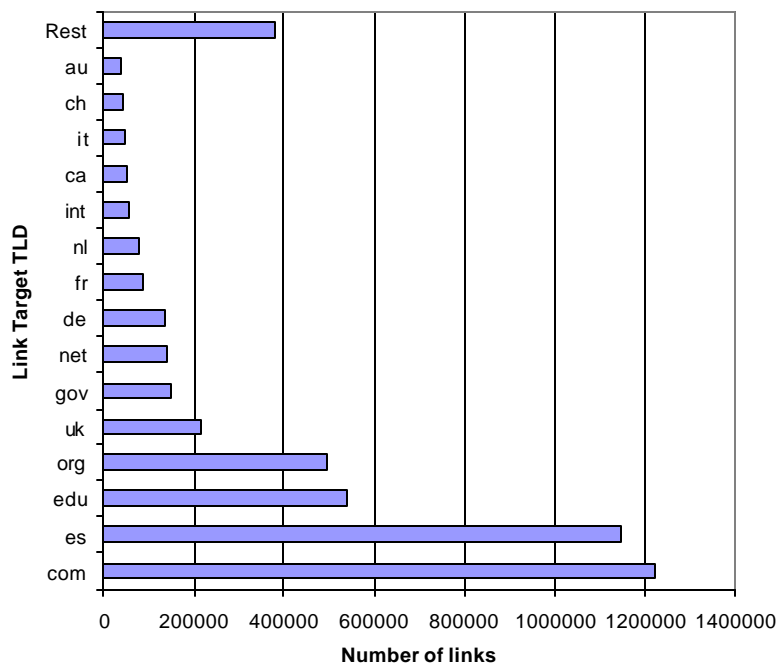


Enlaces a dominios externos

La figura 6 muestra la importancia que tiene el dominio .com para España, atrayendo más enlaces que cualquier otra fuente española. Las universidades americanas (dominio .edu) y el dominio org figuran a continuación. Con relación a otros países, es interesante señalar que los que figuran prominentemente son europeos o angloparlantes, ninguno de los cuales utiliza el español. Interpretamos la distribución de enlaces a gran variedad de destinos internacionales como muy saludable. La ausencia de países hispanohablantes evidencia el multilingüismo de los autores Web españoles. Este

resultado se obtuvo con el modelo página, aunque el modelo dominio no difirió en más de un 2% en cualquier categoría, lo que es un resultado estadísticamente robusto.

Figura 6. Gráfico donde se muestran los dominios (TLD: top level domain) de países e internacionales enlazados desde cualquiera de las sedes universitarias españolas (excluidos enlaces a páginas de la misma universidad).



Páginas preferidas

La tabla 2 suministra información sobre las páginas que atrajeron la mayoría de los enlaces. Esta lista difiere de una similar del Reino Unido, en la cual solo se consideraron enlaces nacionales (29) aunque hay una página de este país que aparece en ambos casos. Ambas listas contienen una página de enlaces a todas las universidades dentro de sus respectivos países.

Tabla 2. Las 25 páginas más enlazadas (inlinks) desde sedes web de universidades españolas.

Enlaces	Dirección	Comentarios
10713	www.mcu.es/bases/spa/isbn/ISBN.html	Agencia Española del ISBN. Enlazada desde muchas bibliotecas universitarias para recuperar información bibliográfica
4937	www.google.com	Google, el motor de búsqueda más popular en España
3335	www.zope.org/credits	Zope es un servidor "open source" que es citado automáticamente.
3274	squishdot.org	Squishdot: The Open Source Discussion Forum Programas relacionados con Zope
2835	www.celcat.com/webpub.html	CELCAT Web Publisher. Utilizada en un gran número de páginas.
2509	www.mfom.es	Ministerio de Fomento. Enlazado desde el Catálogo Español de Buenas Prácticas.

2087	cubic.bioc.columbia.edu/eva	Proyecto EVA de análisis de estructura de proteínas en el que participan universidades españolas que tienen un "mirror"
2026	www.elpais.es	Periódico El País
1713	www.csic.es	Consejo Superior de Investigaciones Científicas
1422	www.boe.es	Boletín Oficial del Estado
1293	www.yahoo.com	El Directorio más popular en España
1255	www.mfom.es/vivienda	Enlazado desde el Catálogo Español de Buenas Prácticas.
1185	www.apple.es	Apple España.
943	cbl.leeds.ac.uk/nikos/personal.html	LaTeX2HTML. Citado desde las páginas transformadas utilizando el programa
902	s12.sitemeter.com/stats.asp?site=s12apardo	Contador de visitas instalado en muchas páginas
833	www.bestpractices.org	Enlazado desde el Catálogo Español de Buenas Prácticas
798	www.lotus.com	IBM Lotus Software
793	www.abc.es	Periódico ABC
788	www-dsed.llnl.gov/files/programs/unix/latex2html/manual	Conversor LATEX to HTML. Citado desde las páginas transformadas utilizando el programa
735	www.rediris.es/recursos/centros/univ.es.html	Listado completo de universidades en la red de investigación española.
648	www.mcyt.es	Ministerio de Ciencia y Tecnología
617	validator.w3.org/check/referer	Comprueba sintaxis HTML en páginas Web
592	www.uab.es	Universitat Autònoma de Barcelona
587	www.bne.es	Biblioteca Nacional de España
578	www.mec.es	Ministerio de Educación y Cultura

En la tabla se puede observar que dominan la lista las páginas que son el objetivo de enlaces generados automáticamente. Sería más útil si las páginas mayormente seleccionadas fueran aquellas con un resultado de enlaces múltiples procedentes de varias fuentes diferentes, más que un subproducto accidental de la decisión de un individuo sobre el diseño de las páginas web. A pesar de esto, es interesante señalar que algunos organismos gubernamentales están bien representados así como instituciones importantes como la Biblioteca Nacional o el CSIC. Las páginas de informática y prensa, posiblemente debido a los enlaces generados automáticamente aparecen también destacadas.

Conclusión

La utilización por parte del sistema educativo español de la Web parece disfrutar de un razonable buen estado de salud, tanto en términos del tamaño de las sedes web como por el grado de interconexión entre las universidades que demuestra la reutilización de fuentes, aunque el uso de la Web presenta cierto retraso comparándolo con países similares de todo el mundo. El multilingüismo internacional de los enlaces web seleccionados muestra una orientación externa excelente y una perspectiva internacional por parte de los autores de los Webs.

Un hecho poco corriente de los resultados, no identificado en anteriores investigaciones (1) es que algunos organismos gubernamentales fueron identificados dentro del grupo de los 25 principales enlaces seleccionados. Sería interesante ver si este fenómeno se

replicaría en estudios de otros países o si la estrecha conexión gobierno-educación es un fenómeno español.

Agradecimientos

Este trabajo forma parte de las contribuciones del proyecto europeo WISER (www.webindicators.org) financiado por el V Programa Marco de I+D de la Unión Europea (INRP-2001-10002). Los autores agradecen el trabajo constante de todo el equipo del Laboratorio de Internet y especialmente el de Esther Ruiz.

Bibliografía

1. Leydesdorff, L.; Curran, M. Mapping university-industry-government relations on the Internet: the construction of indicators for a knowledge-based economy, *Cybermetrics*, 2000, vol. 4(1), paper 2.
<<http://www.cindoc.csic.es/cybermetrics/articles/v4i1p2.html>>
2. Middleton, I.; McConnell, M.; Davidson, G. Presenting a model for the structure and content of a university World Wide Web site. *Journal of Information Science*, 1999, vol. 25(3), 219-227.
3. Azcona, A.; Ortega, V. El uso de Internet en la Universidad Española. II Congreso Internacional de la Lengua. Valladolid, 16-19 de Octubre de 2001.
<cvc.cervantes.es/obref/congresos/valladolid>
4. Wilkinson, D.; Harries, G.; Thelwall, M.; Price, E. Motivations for academic Web site interlinking: Evidence for the Web as a novel source of information on informal scholarly communication, *Journal of Information Science*, 2003, vol. 29(1), 59-66.
5. Thelwall, M.; Binns, R.; Harries, G.; Page-Kennedy, T.; Price E.; Wilkinson, D. European Union Associated University Websites. *Scientometrics*, 2002, vol. 53(1), 95-111.
6. PNECU. Plan Nacional de Evaluación de la Calidad de las Universidades. Informe Global 1996-2000. Madrid: Consejo de Coordinación Universitaria, MEC. 2002.
<www.mec.es/consejou/calidad/Informes_archivos/Informe_global_PNECU.zip>
7. Bricall, J. M. Informe Universidad 2000. Madrid: CRUE. 2002.
<<http://www.crue.org/informeuniv2000.htm>>
7. Telefónica. La Sociedad de la Información en España 2002. Presente y perspectivas. Madrid: Telefónica, 2002.
<<http://www.telefonica.es/sociedaddelainformacion/espana2002/index.html>>
9. Larson, R. R. Bibliometrics of the World Wide Web: an exploratory analysis of the intellectual structure of cyberspace, *Proceedings of the AISS 59th annual meeting*. 1996
10. Almind, T. C.; Ingwersen, P. Informetric analyses on the world wide web: methodological approaches to 'webometrics'. *Journal of Documentation*, 1997, vol. 53(4), 404-426.
11. Rodríguez Gairín, J. M.. Valorando el impacto de la información en Internet: AltaVista, el "Citation Index" de la Red, *Revista Española de Documentación Científica*, 1997, vol. 20:175-181.
<<http://www.kronosdoc.com/publicacions/altavis.htm>>
12. Rousseau, R. Situations: an exploratory study, *Cybermetrics*, 1997, vol. 1(1), Paper 1.
<www.cindoc.csic.es/cybermetrics/articles/v1i1p1.html>

13. Aguillo, I. F. STM information on the Web and the development of new Internet R&D databases and indicators. In: *Online Information 98: Proceedings*. Learned Information, 1998. 239-243.
14. Cronin, B.; Snyder, H.W.; Rosenbaum, H.; Martinson, A.; Callahan, E. Invoked on the web. *Journal of the American Society for Information Science*, 1998, vol. 49(14), 1319-1328.
15. Ingwersen, P. The calculation of Web Impact Factors. *Journal of Documentation*, 1998, vol. 54(2), 236-243.
16. Cronin, B. Bibliometrics and Beyond: Some thoughts on web-based citation analysis. *Journal of Information Science*, 2001, vol. 27(1), 1-7.
17. Davenport, E; Cronin, B. The citation network as a prototype for representing trust in virtual environments. In: Cronin, B. & Atkins, H. B. (eds.). *The web of knowledge: a festschrift in honor of Eugene Garfield*. Metford, NJ: Information Today Inc. ASIS Monograph Series, 2000, p. 517-534.
18. Borgman, C.; Furner, J. Scholarly communication and bibliometrics. In: Cronin, B. (ed.), *Annual Review of Information Science and Technology 36*, Medford, NJ: Information Today Inc., 2002, p. 3-72.
19. Hernández-Borges, A. A.; Macías-Cervi, P.; Gaspar-Guardado, M. A.; Torres-Álvarez de Arcaya, M. L.; Ruiz-Rabaza, A.; Jiménez-Sosa, A. Can Examination of WWW Usage Statistics and other Indirect Quality Indicators Distinguish the Relative Quality of Medical Web Sites? *Journal of Medical Internet Research*, 1999, vol. 1(1).
<<http://www.jmir.org/1999/1/e1/index.htm>>
20. Smith, A. G. A tale of two web spaces: comparing sites using Web Impact Factors, *Journal of Documentation*, 1999, vol. 55(5), 577-592.
21. Harnad, S.; Carr, L. Integrating, navigating, and analysing open eprint archives through open citation linking (the OpCit project). *Current Science*, 2000, vol. 79(5), 629-638.
22. Harter, S. P.; Ford, C. E. Web-based analyses of e-journal impact: approaches, problems, and issues. *Journal of the American Society for Information Science*, 2000, vol. 51(13), 1159-1176.
23. Thelwall, M. Extracting macroscopic information from web links, *Journal of the American Society for Information Science and Technology*, 2001, vol. 52 (13), 1157-1168.
24. Thomas, O.; Willett, P. Webometric analysis of departments of Librarianship and information science. *Journal of Information Science*, 2000, vol. 26(6), 421-428.
25. Chu, H.; He, S.; Thelwall, M. Library and Information Science Schools in Canada and USA: A Webometric Perspective. *Journal of Education for Library and Information Science*, 2002, vol. 43(2), 110-125.
26. Thelwall, M. Conceptualizing documentation on the Web: an evaluation of different heuristic-based models for counting links between university web sites, *Journal of the American Society for Information Science and Technology*, 2002, vol. 53(12), 995-1005.
27. Thelwall, M.; Tang, R. Disciplinary and linguistic considerations for academic Web linking: An exploratory hyperlink mediated study with Mainland China and Taiwan, University of Wolverhampton. 2002.
28. Smith, A. & Thelwall, M. Web Impact Factors for Australasian Universities, *Scientometrics*, 2002, vol. 54(3), 363-380.

29. Thelwall, M. The top 100 linked pages on UK university Web sites: high inlink counts are not usually directly associated with quality scholarly content, *Journal of Information Science*, 2002, vol. 28(6), 485-493.
30. Snyder, H.; Rosenbaum, H. Can search engines be used for web-link analysis? A critical review. *Journal of Documentation*, 1999, vol. 55(4), 375-384.
31. Rousseau, R. Daily time series of common single word searches in AltaVista and NorthernLight, *Cybermetrics*, 1999, vol. 2/3, paper 2
<<http://www.cindoc.csic.es/cybermetrics/articles/v2i1p2.html>>
32. Bar-Ilan, J. Search Engine Results over Time - A Case Study on Search Engine Stability. *Cybermetrics*, 1999, vol. 2/3, paper 1.
<<http://www.cindoc.csic.es/cybermetrics/articles/v2i1p1.html>>
33. Thelwall, M. The responsiveness of search engine indexes, *Cybermetrics*, 2001, vol. 5(1), paper 1.
<<http://www.cindoc.csic.es/cybermetrics/articles/v5i1p1.html>>
34. Vaughan, L.; Thelwall, M. Scholarly use of the Web: What are the key inducers of links to journal Web sites? *Journal of the American Society for Information Science and Technology*, 2003, vol. 54(1), 29-38.
35. Thelwall, M. A web crawler design for data mining, *Journal of Information Science*, 2001, vol. 27(5), 319-325.
36. Aguillo, I. F. EICSTES. Monographs in Official Statistics. EPROS. Progress Report in Statistics 1999-2000. Pages 87-93. EUROSTAT. Luxemburgo. 2001.
37. Thelwall, M.; Harries, G. The Connection between the Research of a University and Counts of Links to its Web Pages: An Investigation Based Upon a Classification of the Relationships of Pages to the Research of the Host University. *Journal of the American Society for Information Science and Technology*, 2003, vol. 54(7), 594-602.
38. Thelwall, M. Methodologies for Crawler Based Web Surveys, *Internet Research: Electronic Networking and Applications*, 2002, vol. 12(2), 124-138.
39. Lawrence, S.; Giles, C. L. Accessibility of information on the web. *Nature*, 1999, vol. 400, 107-109.
40. Aguillo, I. F. (2000). Internet invisible o infranet: Definición, clasificación y evaluación. *Actas de las VII Jornadas Españolas de Documentación Automatizada*. Bilbao, 19-21 Octubre 2000, p. 249-269. FESABID 2000.

Apéndice: 64 dominios web universitarios explorados

Dirección Web	Nombre de la Universidad
http://www.ceu.es/	Universidad de San Pablo CEU
http://www.deusto.es/	Universidad de Deusto
http://www.ehu.es/	Universidad del País Vasco
http://www.muni.es/	Mondragon Unibertsitatea
http://www.ua.es/	Universidad de Alicante
http://www.uab.es/	Universitat Autònoma de Barcelona
http://www.uah.es/	Universidad de Alcalá de Henares
http://www.ualm.es/	Universidad de Almería
http://www.uam.es/	Universidad Autónoma de Madrid
http://www.uax.es/	Universidad Alfonso X El Sabio
http://www.ub.es/	Universitat de Barcelona
http://www.ubu.es/	Universidad de Burgos
http://www.uc3m.es/	Universidad Carlos III de Madrid
http://www.uca.es/	Universidad de Cádiz
http://www.ucjc.es/	Universidad Camilo José Cela
http://www.uclm.es/	Universidad de Castilla La Mancha
http://www.ucm.es/	Universidad Complutense de Madrid
http://www.uco.es/	Universidad de Córdoba
http://www.udc.es/	Universidad de La Coruña
http://www.udg.es/	Universidad de Gerona
http://www.udl.es/	Universitat de Lleida
http://www.uem.es/	Universidad Europea de Madrid
http://www.ugr.es/	Universidad de Granada
http://www.uhu.es/	Universidad de Huelva
http://www.uia.es/	Universidad Internacional de Andalucía
http://www.uib.es/	Universitat de Les Illes Balears
http://www.uimp.es/	Universidad Internacional Menéndez Pelayo
http://www.ujaen.es/	Universidad de Jaén
http://www.uji.es/	Universitat Jaume I de Castellón
http://www.ull.es/	Universidad de La Laguna
http://www.ulpgc.es/	Universidad de Las Palmas de Gran Canaria
http://www.um.es/	Universidad de Murcia
http://www.uma.es/	Universidad de Málaga
http://www.umh.es/	Universidad Miguel Hernández de Elche
http://www.unav.es/	Universidad de Navarra
http://www.unavarra.es/	Universidad Pública de Navarra
http://www.uned.es/	Universidad Nacional de Educación a Distancia
http://www.unex.es/	Universidad de Extremadura
http://www.unican.es/	Universidad de Cantabria
http://www.unileon.es/	Universidad de León
http://www.uniovi.es/	Universidad de Oviedo
http://www.unirioja.es/	Universidad de La Rioja
http://www.unizar.es/	Universidad de Zaragoza
http://www.unnet.es/	Universidad Antonio de Nebrija
http://www.uoc.es/	Universitat Oberta de Catalunya
http://www.upc.es/	Universidad Politécnica de Cataluña

http://www.upco.es/	Universidad Pontificia Comillas
http://www.upct.es/	Universidad Politécnica de Cartagena
http://www.upf.es/	Universitat Pompeu Fabra
http://www.upm.es/	Universidad Politécnica de Madrid
http://www.upo.es/	Universidad Pablo de Olavide
http://www.upsa.es/	Universidad Pontificia de Salamanca
http://www.upv.es/	Universidad Politécnica de Valencia
http://www.urjc.es/	Universidad Rey Juan Carlos
http://www.url.es/	Universitat Ramón Llull
http://www.urv.es/	Universitat Rovira I Virgili Tarragona
http://www.us.es/	Universidad de Sevilla
http://www.usal.es/	Universidad de Salamanca
http://www.usc.es/	Universidad de Santiago de Compostela
http://www.usek.es/	Universidad Internacional SEK
http://www.uv.es/	Universidad de Valencia
http://www.uva.es/	Universidad de Valladolid
http://www.uvic.es/	Universitat de Vic
http://www.uvigo.es/	Universidad de Vigo