

Webometrics pone orden en las universidades

Por Mari-Carmen Marcos

Definición y objetivos

WEBOMETRICS ES EL RESULTADO de un estudio llevado a cabo por el Laboratorio de Internet del Cindoc (Csic), un centro que se define como “observatorio de las actividades de investigación científica y académica, y su publicación en la Web”. Sus investigadores realizan estudios cuantitativos de la información publicada en la Web, a lo que se viene llamando Cibermetría, y más concretamente les interesa estudiar la comunicación científica.

<http://internetlab.cindoc.csic.es>

Tradicionalmente la bibliometría ha estudiado la comunicación científica por medio de la cuantificación de las citas entre artículos. En cambio, en este caso el objeto de estudio es más amplio, pues se trata de analizar las sedes web de instituciones académicas de todo el mundo, mayoritariamente universidades pero también otros centros de investigación (escuelas de negocios, conservatorios superiores o seminarios teológicos, entre otros), en total 13.000 instituciones, de las que 9.500 son universidades.

Las maneras de cuantificar la Web son variadas, y en este estudio se consideran fundamentalmente dos tipos de indicadores:

—volumen de información publicada, y

—número de enlaces que otras sedes web tienen hacia las que son objeto de análisis.

El objetivo primero y más evidente de *Webometrics* es ofrecer un ranking de instituciones que sirva como visión panorámica del estado de este tipo de webs, pero su verdadera y última finalidad es convertirse en una herramienta útil para la evaluación de la investigación y la comunicación en el ámbito académico. Es importante señalar que la posición de cada universidad o centro en el ranking es un reflejo de cómo se sitúa en *comparación* con las webs de otras universidades, y nunca un indicador de la excelencia académica de la universidad en sí.

Sus resultados se actualizan dos veces al año: en enero y en julio. Los primeros resultados publicados datan de julio de 2005, si bien el pasado mes de noviembre se publicó

de forma excepcional un nuevo ranking debido a que algunos de los más importantes motores habían hecho grandes cambios.

Metodología

En primer lugar, y puesto que todo el análisis se realiza sobre los dominios de las webs, es imprescindible que éstos estén claramente definidos y sean perfectamente identificables, pues de lo contrario podrían considerarse páginas que no correspondieran a ellos u omitir algunas que sí lo son. El hecho de necesitar identificar los dominios supone en ocasiones una gran dificultad, pues muchas universidades pequeñas de países en vías de desarrollo todavía no disponen de uno propio, o incluso no tienen web. El caso opuesto se da cuando una institución cuenta con varios dominios principales, o bien en situaciones de fusión de dos o más universidades que deciden mantener sólo uno o ambos. Casos más extraños pero también posibles son aquellos en los que una universidad comparte dominio con otra institución, como la *Universidad de Helsinki*, que lo comparte con el ayuntamiento, lo

que repercute en un mejor posicionamiento en el ranking final, algo que no corresponde con la realidad.

Como decíamos, la metodología de *Webometrics* se basa en el estudio de enlaces, una adaptación del conocido “factor de impacto” que utiliza el *Institute for Scientific Information*, ahora *Thomson Scientific*, tomando los datos de citaciones. En este caso, al tratarse de sitios web y no de artículos publicados en revistas, se han adaptado los indicadores para tener en cuenta aspectos propios de este medio. El factor de impacto específico para la web fue propuesto por primera vez por **Almind** e **Ingwersen**. La cantidad obtenida resulta de dividir el número de enlaces externos que recibe una sede o un dominio web por el número de páginas que componen dicha sede o dominio, de manera que se tienen en cuenta los dos aspectos principales: la visibilidad (a mayor citación mayor visibilidad) y el tamaño (medido en número de páginas que lo componen).

Ahora bien, la importancia de ambos factores no es la misma cuando se trata de sedes web académicas, que es el caso que nos concierne, pues hay determinados ficheros que contienen información más rica y que no ocuparán necesariamente muchas páginas web, por ejemplo los ficheros pdf, doc o ps, que se utilizan para publicar artículos científicos, o los ficheros ppt usados para las presentaciones académicas. Se trata lógicamente de documentos con mayor riqueza que otros. La metodología de análisis de *Webometrics* ha tenido en cuenta esto y da mayor peso a la visibilidad que al tamaño. Ésta es la fórmula



Presentación del sitio web Webometrics

mula de pesos aplicada para conseguir la posición en el ranking:

$$\text{Posición} = 2 * \text{pos}(\text{tamaño}) + 4 * \text{pos}(\text{visibilidad}) + 1 * \text{pos}(\text{ficheros ricos})$$

Para obtener los datos se hace uso de robots que rastrean la Web,

de los cuales algunos son de desarrollo propio y otros son los que ponen a disposición de todos nosotros los APIs de *Google* y otros buscadores. Hasta el momento los motores utilizados han sido *Google*, *Yahoo! Search*, *MSN Search* y *Teoma*. Es posible que en un futuro se in-



Líneas de trabajo del Internet Lab

- Diseño y desarrollo de un observatorio virtual de la presencia en internet de las instituciones, organizaciones e investigadores involucrados en la actividad científico-técnica con cobertura mundial.
- Puesta en marcha de una serie de indicadores cibernéticos para la descripción cuantitativa, con fines de evaluación, de la comunicación científica en la Web.
- Análisis de métodos automáticos para la identificación, recolección y evaluación de recursos web.
- Adecuación y transferencia de las normas del análisis documental a la descripción formal y sustantiva de sedes web, tanto para el estudio de las mismas como para su incorporación a directorios, portales y otros servicios de valor añadido.
- Evaluación periódica de los motores de búsqueda y otras herramientas de localización y recuperación de información en internet.

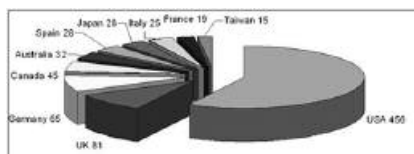
Equipo:

- Begoña Granadino** (directora)
- Isidro F. Aguillo** (coordinador científico)
- José Antonio Prieto** (coordinador de informática)
- Víctor M. Pareja** (asesor técnico)

Colaboradores:

- José Luis Ortega**, **Ana Amieva**, **Teresa Silió**, **Helena Zamora Meca** y **Miguel Atienza de Andrés**.

<http://internetlab.cindoc.csic.es/>



Distribución por países de las 1000 primeras universidades de acuerdo con indicadores cibernéticos

Entrevista a Isidro F. Aguillo, *Webometrics*

En esta época de las nuevas tecnologías, el científico también está viendo cómo su rol cambia ¿qué cambios son los más llamativos?

—Unesco ya reconocía hace varias décadas el papel social del científico, aunque dentro de la comunidad investigadora ese fin sólo se ha aceptado muy recientemente. Los nuevos modos y modelos de comunicación académica y científica a través de internet ayudan notablemente a asumir esa responsabilidad social. El mero hecho de publicar en la Web de forma abierta incrementa su visibilidad y sus actividades, democratiza la ciencia y permite el control ciudadano de una investigación que se financia mayoritariamente con fondos públicos.



En los últimos años estamos viviendo un proceso de evaluación en las universidades para determinar la excelencia de docentes e instituciones. ¿Cómo se traspaesa esta evaluación cualitativa/cuantitativa al ámbito de la Web?

—Hemos de reconocer que la mayoría de los métodos actuales presentan numerosos inconvenientes e importantes sesgos. Sólo procesos muy exhaustivos, costosos en recursos y tiempo pueden garantizar una correcta evaluación. La incorporación de indicadores Web ofrece ventajas indudables que se pueden utilizar de forma complementaria con las técnicas bibliométricas. Destacaríamos la cobertura más amplia de actividades y resultados de investigación, no centrándose únicamente en la producción formal (artículos, informes, patentes, etc.), y la mejor consideración de disciplinas como la tecnología, las ciencias sociales o las humanidades tradicionalmente relegadas en estos estudios.

Si tuviera que enumerar las ventajas que los académicos pueden encontrar para usar la Web como medio de comunicación y difusión ¿qué les diría? ¿y qué reticencias se encontraría?

—La ventaja fundamental es el incremento sustancial de la audiencia potencial, órdenes de magnitud mayor que en cualquier publicación en papel. Esta es una gran virtud tanto para países avanzados como especialmente para los que están en vías de desarrollo pues tanto el acceso como la publicación resultan económicos.

La Web no tiene la rigidez del sistema formal de las revistas científicas y permite la publicación no sólo de todas las fases del proceso de comunicación, sino que además facilita el acceso directo a los datos en bruto, la incorporación de material multimedia y/o interactivo y el establecimiento de foros de discusión.

Las principales reticencias tienen que ver con una mala interpretación del modelo *open access*, que en modo alguno coarta la publicación en revistas de alto factor de impacto, y el temor al abandono del sistema *peer review* (evaluación de los originales por parte de pares o colegas), que sin embargo puede aplicarse independientemente del soporte final de la publicación.

Webometrics nos muestra el resultado del análisis a día de hoy, pero ¿cuáles son las claves para mejorar las posiciones? ¿qué ocurre en las sedes web de los primeros puestos que no ocurre en las demás?

—El viejo lema de “publicar o perecer” es también aplicable aquí, sólo que teniendo en cuenta que los intermediarios clave en la Web son los motores de búsqueda. Se trata de asumir que hay que poner en la Web todo lo que la institución y su personal produzcan, preferiblemente a texto completo y con algún valor añadido. El compromiso real de las universidades anglosajonas con estos principios tiene su reflejo en el excelente rendimiento que se observa en *Webometrics*. Por el contrario, el retraso por parte de las universidades europeas y latinoamericanas está dando lugar a una preocupante “brecha digital”.

Isidro F. Aguillo es coordinador científico del Observatorio de Ciencia y Tecnología en Internet del CINDOC (CSIC) e investigador principal del proyecto *Webometrics*.

isidro@cindoc.csic.es

RANKING MUNDIAL	UNIVERSIDAD	PAÍS	POSICIÓN		FICHEROS RICOS
			TAMAÑO	VISIBILIDAD	
1	MASSACHUSETTS INSTITUTE OF TECHNOLOGY		1	1	8
2	UNIVERSITY OF CALIFORNIA BERKELEY		2	2	2
3	STANFORD UNIVERSITY		4	5	1
4	HARVARD UNIVERSITY		3	3	13
5	CORNELL UNIVERSITY		6	4	12
6	UNIVERSITY OF ILLINOIS URBANA CHAMPAIGN		8	6	3
7	UNIVERSITY OF TEXAS AUSTIN		9	8	7
8	UNIVERSITY OF WASHINGTON		5	12	5
9	UNIVERSITY OF MICHIGAN		11	7	19
10	UNIVERSITY OF WISCONSIN MADISON		10	10	9
11	UNIVERSITY OF MINNESOTA		14	9	6

Primeros puestos del ranking mundial de universidades

corporen otros como *Gigablast*, *Exalead*, *Alexa* y *Google Scholar*.

Un factor que parece no haberse tenido en cuenta es el del dinamismo para la creación de páginas, es decir, los contenidos que se encuentran en bases de datos, pues al no ser páginas estáticas los buscadores no los indexan. En cambio se trata de una información extremadamente valiosa para un estudio sobre comunicación de información científica, ya que pueden existir bases de datos alojadas en los dominios de las universidades que contengan artículos publicados por los profesores, tesis defendidas por los estudiantes, memorias de los proyectos de investigación, etc. Limitar el análisis a la web “visible” (claro está, visible para los motores) es dejar de lado información de la que no sabremos ni cuánta es ni de qué calidad para el objetivo académico de *Webometrics*.

En general, como fuente de información cuantitativa fiable, los motores presentan distintos problemas, pero ésta es la razón de utilizar valores relativos (ranking) en vez de absolutos. De todas formas, posiblemente el impacto del redondeo en *Google* (error mínimo del 5%), la persistencia de páginas desaparecidas, las páginas huérfanas y otras

circunstancias sean todavía más relevantes globalmente que la indización de la web invisible¹.

Resultados

La web de *Webometrics* presenta los resultados obtenidos en el último estudio realizado (a fecha 2 de enero de 2006 los resultados que muestra la web son de noviembre de 2005).

http://www.webometrics.info/index_es.html

El ranking global es el “Ranking cibernético de universidades del mundo”, que recoge las 3.000

universidades que han obtenido mayor puntuación.

También se ofrecen los rankings por zonas geográficas, así como los centros especiales que no son universidades.

En el ranking de Europa se incluyen las 500 universidades mejor puntuadas. Se puede observar que la primera universidad europea, *Cambridge*, aparece en el ranking mundial en la posición número 21, y la segunda europea, *Oxford*, ocupa el puesto 28 de las de todo el mundo. Si observamos el caso español, vemos que la *Complutense de Madrid* es la primera en aparecer en ambos rankings (tiene el puesto 94 dentro de las europeas y el 282 en todo el mundo), a pocos puestos de la *Autònoma de Barcelona* (que ocupa el 98 y el 288 respectivamente).

Para España se ha elaborado una sección exclusiva en la que se presenta el ranking de las 100 universidades mejor puntuadas, y se completa con los datos de la posición que ocupa cada centro en el ranking mundial en cuatro aspectos: posición global, posición en cuanto al tamaño, en cuanto a la visibilidad y en cuanto a cantidad de ficheros ricos. El análisis se expone

Top Europa			
Universidades 1 a 100 de 100			
RANKING CONTINENTAL	UNIVERSIDAD	PAÍS	RANKING MUNDIAL
1	UNIVERSITY OF CAMBRIDGE		21
2	UNIVERSITY OF OXFORD		28
3	SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH		41
4	UNIVERSITY OF EDINBURGH		54
5	UNIVERSITY OF HELSINKI		63
6	LINKÖPING UNIVERSITY		65
7	UNIVERSITY COLLEGE LONDON		68
8	UNIVERSITY OF OSLO		72
9	UNIVERSITY OF LEEDS		73
10	UNIVERSITY OF BOLOGNA		78

además dividido en cuatro zonas geográficas: Norte (31 centros); Sur-Este (38 centros); Madrid (30 centros) y Cataluña (51 centros). Para los centros de investigación que no son universidades (institutos, laboratorios, etc.) hay un ranking aparte.

Latinoamérica muestra un total de 200 universidades de las que la número uno es la de *São Paulo* (109 en el mundial) y la segunda es la *Nacional de México* (149 en el mundial), con una clara mayoría de las brasileñas en los primeros puestos. Se ha preparado una sección con resultados por zonas geográficas: Brasil, Pacto Andino, Cono Sur, España y Portugal (excluidos en el ranking general de Latinoamérica), y América central (incluido el Caribe).

En Asia es Japón quien acapara buena parte de los puestos privilegiados: de las 15 universidades mejor situadas de Asia, 8 son de Japón y 3 de Israel. Singapur, Taiwán, Hong Kong y China tienen las otras 4 en esos primeros 15 puestos.

En cuanto a Oceanía, no es nada despreciable el puesto de sus tres universidades mejor posicionadas: la *Australian National*, la de *Melbourne* y la de *Sydney* están en dentro de las 100 mejores del mundo.

Además de las listas por países, continentes y subzonas geográficas, *Webometrics* aporta otros datos que pueden servir para conocer el universo Web de cada país:

—Número de dominios.

—Número de páginas que hay de cada dominio en cada país (páginas con dominio nacional, .com, .info, etc.).

—Ratio entre el número de páginas web y el número de habitantes.

—Ratio entre el número de páginas web y el número de dominios académicos.

ESPAÑA	MUNDO	UNIVERSIDAD	TAMAÑO	VISIBILIDAD	ACTIVIDADES
1	282	UNIVERSITY COMPLUTENSE MADRID	217	394	136
2	288	AUTONOMOUS UNIVERSITY OF BARCELONA	316	359	150
3	301	UNIVERSITY OF BARCELONA (UB.ES)	249	388	260
4	308	POLYTECHNIC UNIVERSITY OF MADRID	336	342	334
5	327	POLYTECHNIC UNIVERSITY OF CATALONIA	334	369	351
6	333	UNIVERSITY OF VALENCIA	347	431	129
7	371	UNIVERSITY OF GRANADA	532	358	307
8	375	AUTONOMOUS UNIVERSITY OF MADRID	152	573	226
9	385	UNIVERSITY OF ZARAGOZA	440	457	179
10	391	POLYTECHNIC UNIVERSITY OF VALENCIA	304	523	217
11	431	UNIVERSITY OF ALICANTE	303	631	107
12	435	UNIVERSITY OF OVIEDO	268	593	363
13	448	UNIVERSITY OF SEVILLE	248	674	176
14	493	UNIVERSITY OF THE BASQUE COUNTRY	712	460	443
15	515	UNIVERSITY OF THE BALEARIC ISLANDS	465	680	332
16	534	UNIVERSITY OF VALLADOLID	563	635	455
17	535	UNIVERSITY OF NAVARRA	265	744	618
18	537	SPANISH DISTANCE EDUCATION UNIVERSITY	772	543	429
19	585	JALME I UNIVERSITY	609	726	410
20	627	UNIVERSITY OF SANTIAGO DE COMPOSTELA	459	830	582
21	632	UNIVERSITY OF MURCIA	473	875	394
22	692	UNIVERSITY POMPEU FABRA (UPF.ES)	923	633	852
23	723	UNIVERSITY CARLOS III MADRID	722	890	355
24	747	UNIVERSITY OF MALAGA	756	899	485

Primeros puestos del ranking de universidades en España e indicación de su posición en el mundial

—Ratio entre el número de páginas web y el producto interior bruto del país.

Veamos un ejemplo, España:

—Tiene 2.700.000 dominios (puesto 26).

—Sumando todos los dominios obtenemos en total 65.207.000 páginas web (puesto 12).

—Este número da un total de 1.587 páginas por habitante (puesto 35).

—Hay 390.461 páginas por dominio académico (puesto 58).

—La relación entre el número de páginas web de España y su PIB es de 78 (puesto 79).

Un producto derivado del trabajo realizado del ranking y que se

pone a disposición de todos es el catálogo de universidades y de centros de I+D, donde se recogen por continentes y países todos los centros de enseñanza superior y de investigación que se han identificado con una URL propia.

Otros proyectos en la misma línea

Webometrics se inscribe dentro de los estudios estadísticos de comparación de actividades de docencia e investigación. Algunos de estos proyectos de especial interés son:

—*Academic ranking of world universities*, elaborado por la universidad *Shanghai Jiao Tong*.

—Ranking del periódico *The Times* en su suplemento *Higher Education*.

—Informes del *ISI* publicados en la web *In-cites*, realizados a partir de los datos recogidos en su base de datos *Essential Science Indicators*.

Para el futuro

Un estudio como el que hemos presentado nunca acaba; sus datos nunca pueden darse como definitivos, pues la web es un medio cambiante en el que posicionarse depende en gran medida de la voluntad que cada institución tenga por comunicar y difundir a través de ella. En cambio no debe confundirse la *visibilidad* que la web de una universidad tenga, con su *calidad o su excelencia académica*, pues son dos aspectos que no necesariamente están relacionados, aunque quizá en buena parte de estos centros coincida un buen posicionamiento de su web con una alta calidad. Esta relación se podrá conocer más adelante, y si se viera que existe cierta correlación sería debido a que las universidades más presti-

Top Latino América			
Universidades 1 a 100 de 100			
RANKING CONTINENTAL	UNIVERSIDAD	PAÍS	RANKING MUNDIAL
1	UNIVERSITY OF SAO PAULO		109
2	UNIVERSITY NACIONAL AUTONOMA DE MEXICO		149
3	STATE UNIVERSITY OF CAMPINAS		187
4	UNIVERSITY OF CHILE		293
5	TECNOLOGICO DE MONTERREY		413
6	FEDERAL UNIVERSITY OF RIO DE JANEIRO		432
7	UNIVERSITY OF BUENOS AIRES		438
8	FEDERAL UNIVERSITY OF SANTA CATARINA		446
9	FEDERAL UNIVERSITY OF RIO GRANDE DO SUL		475
10	PONTIFICAL CATHOLIC UNIVERSITY OF RIO DE JANEIRO		518

Primeros puestos del ranking de universidades en Latinoamérica

giosas han optado por dar a conocer su investigación por medio de publicaciones en la web.

Webometrics está ahora dando sus primeros frutos y su equipo trabaja para afinar algunos detalles del cálculo de los indicadores y aumentarlos. Llegado el momento podre-

mos conocer los resultados de correlacionar los indicadores de tamaño y visibilidad con otros que midan la excelencia académica y científica de los centros analizados.

Nota

1. Alguien puede argumentar que las páginas incluidas en bases de datos sí son accesibles a los robots. Es un asunto difícil de contestar, debido a la evolución de éstos.

—En principio, las bases de datos con password son internet privada. Al no ser Web pública *Webometrics* no hace estudios con ellas.

—El conjunto de las bases de datos con pasarela de acceso libre pero que no son indizadas por los motores de búsqueda se llama internet o web invisible. Por ejemplo, en algunos casos hacen rellenar un formulario.

—Pero si tienen un menú desplegable de apoyo (filtro de provincias, especialidades, etc.) entonces los robots pueden volcar una parte considerable de ellas. Por ejemplo, *Google* tiene indexados 10 de los 16 millones de registros de *PubMed*.

—Si tienen páginas «estáticas» que enlazan a resultados de búsqueda, los motores también lo fichan. Así parece funcionar con *Dialnet*, que está indexado al completo.

—También pueden crearse páginas caché como la biblioteca de la *Universidad de Navarra*, lo que hace que en la Red haya 500.000 páginas de registros bibliográficos de esta universidad.

El impacto webométrico de todo ello puede ser grande, por lo que esa es la razón por la que cuando en *Webometrics* calculan el tamaño de una sede eliminan los valores máximo y mínimo de los motores de búsqueda (generalmente *Google* es el que da el superior). Es decir que el impacto de la internet invisible en cibermetría tiene carácter puntual, depende del momento y de muchas circunstancias.

Top Asia			
Universidades 1 a 100 de 100			
RANKING CONTINENTAL	UNIVERSIDAD	PAÍS	RANKING MUNDIAL
1	TOKYO UNIVERSITY		83
2	NATIONAL UNIVERSITY OF SINGAPORE		121
3	KEIO UNIVERSITY		132
4	HEBREW UNIVERSITY OF JERUSALEM		148
5	NATIONAL TAIWAN UNIVERSITY		170
6	TECHNION ISRAEL INSTITUTE OF TECHNOLOGY		183
7	KYOTO UNIVERSITY		190
8	NAGOYA UNIVERSITY		192
9	CHINESE UNIVERSITY OF HONG KONG		211
10	PEKING UNIVERSITY		213
11	TOKYO INSTITUTE OF TECHNOLOGY		247
12	OSAKA UNIVERSITY		252
13	UNIVERSITY OF TSUKUBA		276
14	TEL AVIV UNIVERSITY		280
15	TOHOKU UNIVERSITY		289

Primeros puestos del ranking de universidades en Asia