

Presencia y visibilidad web de las universidades públicas españolas

Enrique Orduña-Malea*, Jorge Serrano-Cobos**, José-Antonio Ontalba-Ruipérez*, Nuria Lloret-Romero*

Resumen: Se estudia la evolución del tamaño y visibilidad de los dominios web de las universidades públicas españolas desde enero hasta junio de 2009 en función de diversos buscadores web (*Google, Yahoo!, Live/Bing* y *Exalead*). Asimismo, se proponen el *factor de representatividad relativa media en tamaño (Rs)* y el *factor de representatividad relativa media en visibilidad (Rv)* como indicadores para conocer la importancia de un dominio, en número de documentos y de enlaces externos respectivamente, durante un intervalo temporal, con el propósito de ser aplicados en el diseño y construcción de rankings de universidades basados en técnicas cibernéticas.

Los resultados confirman que las diferencias de tamaño entre los dominios académicos web son muy grandes según el buscador utilizado, por lo que el uso de un único buscador no permite conocer de forma fiable el tamaño real de un dominio web. Además, el empleo de valores combinados a partir de la media obtenida en cada buscador no ofrece resultados fiables, dada la varianza de datos obtenida entre buscadores así como los diferentes índices obtenidos de *Rs*. Las diferencias en visibilidad son menores aunque igualmente significativas.

Se comprueba que los indicadores *Rs* y *Rv* proporcionan información útil y coherente acerca del grado de evolución de las universidades en la Web durante un intervalo temporal determinado. Se obtiene asimismo una correlación positiva entre estos dos indicadores tanto en *Yahoo!* como en *Exalead*, que confirma la relación entre el tamaño documental de un dominio académico web y el número de enlaces que recibe a lo largo del tiempo.

Palabras clave: Universidades públicas españolas, dominios académicos web, tamaño web, visibilidad web, España, ranking de universidades.

Web presence and visibility of Spanish public universities

Abstract: *The evolution of size and visibility of the Spanish public universities websites according to various search engines (Google, Yahoo!, Live/Bing y Exalead) was studied from January to June 2009. Additionally, the article proposes two indicators for understanding the importance of a web domain: the Relative representativeness size factor (Rs) and the Relative representativeness visibility factor (Rv). These indicators, which consider the number of both documents and links, respectively, during a specific*

* Universidad Politécnica de Valencia, Escuela Técnica Superior de Ingeniería Informática. Valencia, España. Correo-e: enorma@upv.es, joonrui@upv.es, nlloret@upvnet.upv.es.

** MASmedios.com, Valencia, España. Correo-e: jorge@masmedios.com.

Recibido: 16-7-09; 2.ª versión 24-9-09; aceptado: 29-9-09.

interval of time are intended to be applied in the design and construction of university rankings based on cybermetric techniques.

The results confirm that the size differences among academic web domains vary significantly depending on the search engine used; therefore the use of a single web browser cannot supply reliable information about the actual size of the web domain. Moreover, the use of combined values from the mean obtained from each search engine does not offer reliable results, given the variance of data obtained from the different search engines, as well as the index differences of Rs. The differences concerning visibility were smaller, but significant nonetheless.

Rs and Rv indicators were found to provide useful and consistent information about the level of development of universities on the Web during a given time interval. There was also a positive correlation between these two indicators on both Yahoo! and Exalead, confirming the relationship between the number of documents of an academic web domain and the number of links it receives over time.

Keywords: *Spanish public universities, academic web domains, web size, web visibility, Spain, University ranking.*

1. Introducción

La inquietud y necesidad de realizar comparaciones entre instituciones universitarias, pese a ser un tema de actualidad e impacto mediático, no son algo propio del siglo XXI. Algunos autores señalan el inicio de estas actividades en Estados Unidos a través de ciertos informes estadísticos anuales que el *United States Bureau of Education* comienza a publicar hacia 1870 (Stuart, 1995). No es hasta 1983, con la primera publicación del *U.S News & World Report (USNWR)*¹, cuando nacen los rankings en la forma en que los entendemos hoy día.

Pese a que los procedimientos, indicadores e incluso los vehículos documentales para publicar rankings (Web, publicación comercial impresa, informes, artículos científicos, etc.) han variado con el paso de los años, éstos nunca han estado exentos de polémica, debido fundamentalmente a debilidades en los procedimientos técnicos y metodológicos, comparaciones entre universidades con misiones y objetivos diferentes y desconocimiento de la verdadera utilidad de los resultados e impacto real en los usuarios (Hazelkorn, 2007).

La idea detrás de todas las iniciativas de ranking siempre ha sido la de capturar la calidad universitaria a partir de una serie de indicadores que, combinados y ponderados de alguna determinada forma, permitiesen clasificar las universidades en un determinado orden, en función de la idea de calidad expresada a partir de los indicadores usados. De esta forma, existen rankings que prestan más atención a determinados tipos de indicadores que a otros; por ejemplo, *USNWR* da importancia a las aptitudes previas de los alumnos mientras que el ranking de la *Universidad de Leiden*² se centra en la investigación académica, a través de indicadores de naturaleza exclusivamente bibliométrica. De todas las tipologías de indicadores existentes, Usher y Savino (2006) proponen la más completa en la actualidad³.

Con la creciente globalización de los mercados, el progresivo aumento de la movilidad internacional de estudiantes y la cada vez más común idea de la educación como un bien económico (Altbach, 2006), el concepto de ranking como herramienta de apoyo a la toma de decisiones crece de forma imparable.

Así, a principios del siglo XXI se entra en una nueva fase gracias a tres importantes hitos: la primera publicación de un ranking de cobertura mundial⁴ (2003), el nacimiento del *IREG*⁵ (2004) así como la constitución de los principios de Berlín (2006). A partir de entonces, el nacimiento de nuevos rankings mundiales y nacionales se sucede de forma continua.

Entre éstos, destaca el lanzamiento en 2004 del *Ranking Web of World Universities* (RWWU)⁶, gracias al trabajo de Isidro F. Aguillo y otros (2008). Este proyecto constituye la primera iniciativa seria centrada en realizar un ranking de universidades a partir de la producción documental web.

Con posterioridad se han publicado otros rankings universitarios, mundiales o nacionales, basados también en indicadores cibernéticos, entre los que destaca el *Web Popularity Ranking*⁷, el *Ranking Universitario de transparencia Web*⁸, el *Ranking Colleges using Google and OSS*⁹, iniciativa personal de Mike Tung, y la propuesta del *G-Factor*¹⁰ por parte de Peter Hirst.

Pese a ciertas críticas acerca de si los métodos cibernéticos son adecuados para medir o calibrar la producción documental de una universidad, subyace una idea de interés documental: la posibilidad de analizar y comparar una serie de instituciones a partir de la producción documental web que éstas generan y consumen (Thelwall, 2001; Aguillo y otros, 2006).

Por tanto, más allá de medir la «calidad», que depende del enfoque bajo el que se estudie, se trata de conocer la presencia (tamaño) y visibilidad (impacto) de una universidad a través de su dominio web, como indicadores de lo bien o mal que ésta se comunica con su comunidad (debido esto en muchos casos a malas prácticas web¹¹). No obstante, los datos que aporta el RWWU tienen la particularidad de:

- a) Mostrar los diferentes parámetros utilizados ya combinados y normalizados, hecho que impide:
 - cuantificar las diferencias según los distintos buscadores usados, y
 - estudiar la validez o no del uso de indicadores combinados a partir del resultado de varios buscadores.

Este punto es de vital importancia, pues el uso de buscadores comerciales ha provocado diversas críticas hacia los métodos cibernéticos debido a problemas de escasa cobertura (Lawrence y Giles, 1999), falta de estabilidad o precisión principalmente (Bar-Ilan, 1999; 2002; 2003) y algunos otros sesgos (Aguillo y Granadino, 2006). Una exagerada diferencia entre los resultados de diferentes buscadores podría invalidar la construcción de un indicador combinado.

- b) Ofrecer una foto fija (2 ediciones, en enero y julio) de los dominios académicos web:
- Se intuye por tanto la necesidad, dentro del dinamismo de la Web, de observar cómo es la evolución de los datos para obtener información más exacta acerca de posibles tendencias.
 - Además, es de interés la propuesta de algún tipo de indicador que mida el grado de dicha evolución, tanto en tamaño como impacto, de los dominios académicos web.

El problema relacionado con el dinamismo de la Web (Brewington, 2000) entronca directamente con los estudios longitudinales y/o evolutivos de la Web, cuya importancia ya ha sido manifestada por diversos autores (Rousseau, 1999). Este campo se basa en el estudio de la perdurabilidad y estabilidad de las páginas web a lo largo del tiempo así como de las variables que las controlan.

En ese sentido, las páginas web pueden mantenerse estables, modificarse, desaparecer y reaparecer (intermitencia), modificadas o no. Por esta razón, los indicadores cibernéticos utilizados en un estudio longitudinal (cuyo objeto es el análisis de los recursos de un dominio web académico con el fin de elaborar un ranking de universidades) deben basar sus mediciones precisamente en la cuantificación de los cambios producidos en los recursos albergados dentro de este dominio web durante el intervalo de estudio.

El problema estriba pues en conocer qué cambios se deben medir y qué factores afectan o favorecen dichos cambios, de forma que ayuden a explicarlos. Además, las modificaciones pueden producirse a diferentes niveles:

- a) Cambios internos de contenido en un recurso que perdura en el tiempo: se cuantifican a través del peso del recurso en bytes.
- b) Cambios internos en la estructura de enlaces de un recurso que perdura en el tiempo: se cuantifican a través de los enlaces que recibe o genera un documento.
- c) Cambios debidos a la publicación o desaparición completa de recursos: se cuantifican mediante el conteo del número de recursos, de diferente formato, alojados dentro de un dominio académico web, y detectados a través de una consulta a un buscador web.

Todos estos cambios se consideran indicadores imperfectos (Koehler, 2002), pues indican un cambio cuantitativo, pero no se captura en ningún caso la calidad o la naturaleza subjetiva del cambio.

La medición del peso del recurso permite detectar modificaciones sin que tenga que desaparecer el documento, aunque existe la posibilidad de que se realicen cambios en el significado del documento, pero no en su tamaño. Además, este indicador resulta difícil de obtener y gestionar para amplios conjuntos de recursos web.

Respecto al conteo total, tanto de enlaces como de recursos, éstos tienen la ventaja de ser relativamente sencillos de obtener, aunque se pueden producir simplemente por cambios en las prestaciones de los buscadores y no por cambios reales.

No obstante, dada su sencillez, la mayoría de rankings web opta por la medición del número total de recursos publicados (*size*) y de enlaces entrantes (*visibility*) como las magnitudes más importantes a cuantificar en un dominio académico web, con el propósito de confeccionar rankings académicos.

Respecto a las variables que pueden influir en la mayor o menor estabilidad de las páginas web a lo largo del tiempo (variables de perdurabilidad), destacan las siguientes:

a) *Dominio web*

Los trabajos de Koehler (1999; 2002; 2004), Cho y García-Molina (2003) y Fetterly y otros (2003) han encontrado evidencias de una mayor persistencia de los dominios *.edu*, respecto del *.com*. Dado el gran número de universidades con dominio *.edu* (sobre todo en Norteamérica), esto podría indicar una mayor persistencia de los dominios académicos estadounidenses.

b) *Contenido*

La vida media de los recursos web en diferentes disciplinas difieren (Koehler (2004). Bar-Ilan y Peritz (2004) detectan asimismo que las páginas web generales son menos estables que los documentos alojados en bibliotecas digitales.

Estos mismos autores se preguntan igualmente si existen temáticas más estables que otras, o si es la forma de los documentos (artículos, conferencias, noticias, etc.) lo que influencia las ratios de modificación y desaparición de páginas web. De esto se podría deducir que aquellas universidades con repositorios o depósitos de documentos científicos, e incluso enfocados a ciertas áreas temáticas, podrían tener estadísticamente una mayor probabilidad de persistir en el tiempo.

Koehler (2002), basándose en las tasas de cambios detectadas, sugiere por su parte la existencia de dos tipos de contenidos educativos en las webs académicas: publicación académica (a largo plazo) y publicación docente o de clase (a corto plazo).

c) *Páginas dinámicas*

Este punto es debatido por Payne y Thelwall (2007) en su análisis de los dominios académicos del Reino Unido, Australia y Nueva Zelanda. En ese estudio reconocen que la aparente estabilización en el número de páginas estáticas en un dominio puede no deberse a un estancamiento

en su expansión, y que esto mismo podría tener un efecto en universidades con sedes web más antiguas y con tiempo de publicar más páginas estáticas en sus dominios y por tanto de construir una mayor presencia web. Pese a esto, las prestaciones de los motores de búsqueda comerciales impiden cualquier estudio de este tipo en la actualidad.

Otras variables influyentes son las siguientes:

- d) *Tipo*: las páginas de navegación tienden a desaparecer menos que las de contenido.
- e) *Edad*: las páginas tienden a modificarse menos conforme pasa el tiempo.
- f) *Tamaño*: las páginas más grandes cambian más frecuente y profundamente que las más pequeñas (Fetterly y otros, 2003).

El conocimiento de la influencia de estas variables puede ayudar en gran medida a interpretar los resultados obtenidos en cualquier análisis cibernético y, por tanto, de un ranking web de universidades.

Pese a que los análisis cibernéticos de instituciones académicas son ya abundantes, todavía existen pocos trabajos llevados a cabo desde una perspectiva longitudinal (Ortega y otros, 2006; Payne y Thelwall, 2007) y que además cubran el espectro de las universidades españolas, donde se destacan los trabajos de Thelwall y Aguillo (2003), Ortega y Aguillo (2007) y Orduña-Malea y otros (2009), este último relativo a la presencia de las universidades españolas en *Google Scholar*.

En ese sentido, el objetivo principal de este trabajo es el de aportar y proponer soluciones a los problemas anteriormente comentados mediante el análisis de la evolución del tamaño y visibilidad de los dominios académicos web de las universidades públicas españolas, con el propósito de mejorar el uso de indicadores cibernéticos en el diseño y realización de rankings de universidades.

Los objetivos concretos que se plantean son los siguientes:

- Medir el tamaño de los dominios académicos web de las universidades públicas españolas para conocer y contextualizar su presencia en la Web en función de diversos buscadores.
- Estudiar la idoneidad de la utilización de indicadores combinados a partir de la cuantificación los datos de distintos buscadores.
- Proponer el *factor de representatividad relativa media en tamaño (Rs)*, para conocer la importancia de un dominio web en número de documentos durante un intervalo temporal.
- Proponer el *factor de representatividad relativa media en visibilidad (Rv)*, para conocer la importancia de un dominio web en número de enlaces externos durante un intervalo temporal.
- Analizar la visibilidad de los dominios académicos web en función de su tamaño, a través del factor de impacto web (*WebIF*) y de la correlación entre *Rs* y *Rv*.

2. Metodología

Se analizan los dominios académicos web de las universidades públicas españolas^{12, 13}, desde enero hasta junio de 2009, ambos meses incluidos (tabla I).

Los datos se obtienen a través de la consulta a la *API* de cada uno de los dominios (tabla II) y mediante técnicas de *screen scraping* (obtención de los datos de la interfaz manual de los buscadores). Para la toma de datos mensual se realizan varias catas por mes y se obtiene la media de dichas muestras. Esta operación se repite para cada indicador, buscador y dominio analizado. Todos los datos obtenidos se exportan durante la primera semana de julio de 2009 a una hoja de cálculo para ser tratados.

TABLA I
Universidades públicas españolas

Universidades	Código	URL
Universidad Autónoma de Barcelona	UAB	http://www.uab.es/
Universidad Autónoma de Madrid	UAM	http://www.uam.es/
Universidad Carlos III de Madrid	UC3M	http://www.uc3m.es
Universidad Complutense de Madrid	UCM	http://www.ucm.es/
Universidad de Alcalá	UAH	http://www.uab.es/
Universidad de Alicante	UA	http://www.ua.es/
Universidad de Almería	UAL	http://www.ual.es/
Universidad de Barcelona	UB	http://www.ub.edu
Universidad de Burgos	UBU	http://www.ubu.es
Universidad de Cádiz	UCA	http://www.uca.es/
Universidad de Cantabria	UC	http://www.unican.es
Universidad de Castilla-La Mancha	UCLM	http://www.uclm.es/
Universidad de Córdoba	UCO	http://www.uco.es/
Universidad de Extremadura	UNEX	http://www.unex.es/
Universidad de Girona	UDG	http://www.udg.edu/
Universidad de Granada	UGR	http://www.ugr.es
Universidad de Huelva	UHU	http://www.ubu.es/
Universidad de Jaén	UJA	http://www.ujaen.es/
Universidad de La Coruña	UDC	http://www.udc.es
Universidad de La Laguna	ULL	http://www.ull.es
Universidad de La Rioja	UR	http://www.unirioja.es/
Universidad de las Islas Baleares	UIB	http://www.uib.es
Universidad de Las Palmas de Gran Canaria	ULPGC	http://www.ulpgc.es/

TABLA I (continuación)

Universidades	Código	URL
Universidad de León	UL	http://www.unileon.es/
Universidad de Lleida	UDL	http://www.udl.es/
Universidad de Málaga	UMA	http://www.uma.es/
Universidad de Murcia	UM	http://www.um.es/
Universidad de Oviedo	UO	http://www.uniovi.es/
Universidad de Salamanca	USAL	http://www.usal.es
Universidad de Santiago de Compostela	USC	http://www.usc.es/
Universidad de Sevilla	US	http://www.us.es/
Universidad de Valencia	UV	http://www.uv.es
Universidad de Valladolid	UVA	http://www.uva.es/
Universidad de Vigo	UVI	http://www.uvigo.es/
Universidad de Zaragoza	UZ	http://www.unizar.es/
Universidad del País Vasco	EHU	http://www.ehu.es
Universidad Jaume I	UJI	http://www.uji.es/
Universidad Miguel Hernández de Elche	UMH	http://www.umb.es/
Universidad Nacional de Educación a Distancia	UNED	http://www.uned.es
Universidad Pablo de Olavide	UPO	http://www.upo.es/
Universidad Politécnica de Cartagena	UPCT	http://www.upct.es/
Universidad Politécnica de Cataluña	UPC	http://www.upc.es/
Universidad Politécnica de Madrid	UPM	http://www2.upm.es
Universidad Politécnica de Valencia	UPV	http://www.upv.es
Universidad Pública de Navarra	UPNA	http://www.unavarra.es/
Universidad Rey Juan Carlos	URJC	http://www.urjc.es/

Fuente: CRUE.

TABLA II

*Ecuaciones de búsqueda usadas
(ejemplo: Universidad Autónoma de Barcelona)*

Buscador	Indicador	
	Tamaño (Size)	Impacto (Visibility)
Google	site:uab.es	No usado
Yahoo!	site:uab.es	linkdomain:uab.es -site:uab.es
Live/Bing	site:uab.es	No usado
Exalead	site:uab.es	link:uab.es -site:uab.es

Los resultados obtenidos no deben ser entendidos como aproximaciones a la calidad de las universidades, sino como una muestra de lo que los buscadores están mostrando de los dominios académicos, síntoma de la comunicación vía web que las universidades públicas tienen con sus usuarios, y de su representatividad e impacto en diferentes motores de búsqueda.

Se procede a la captura de datos de dos tipos de indicadores: tamaño (*size*) e impacto (*visibility*).

a) Tamaño web (*size*)

Se utilizan los buscadores *Google*, *Yahoo!*, *Live*¹⁴ y *Exalead*, por ser los más implantados y además ofrecer comandos apropiados para el cálculo de este indicador.

Puesto que la visualización conjunta de los datos brutos de todas las universidades no es adecuada, dado que los órdenes de magnitud tanto entre ellas como entre los distintos buscadores son muy diferentes, se decide normalizar los valores de 0 a 100 (práctica habitual, con ligeras diferencias, en otros rankings), mediante un proceso de transformación de cociente (Rocki, 2005):

$$v_{sn} = \frac{x_{sn}}{\sum_{i=1}^N x_{sn}} \cdot 100 \quad [1]$$

v_{sn} = valor normalizado obtenido en tamaño (s) para una universidad (n).

x_{sn} = valor bruto obtenido en tamaño (s) para una universidad (n).

N = Conjunto de universidades estudiadas.

De este modo, la suma del tamaño obtenido de todas las universidades durante un mes se considera igual a 100, y el valor de cada universidad se calcula de forma proporcional.

Tras normalizar resultados, ya no se obtienen evoluciones cuantitativas totales, sino un porcentaje de tamaño proporcional al total obtenido por todas las universidades en un motor de búsqueda determinado cada mes (concepto denominado «representatividad relativa» en este trabajo).

Posteriormente, se calcula la media del v_{sn} calculado cada mes y se obtiene un valor, también entre 0 y 100, cuya denominación se propone como «factor de representatividad relativa media en tamaño (R_s)». Este factor puede calcularse para cualquier conjunto de sedes seleccionadas y cualquier período de tiempo.

$$R_s = \frac{\sum_{i=1}^M v_{sn}}{M} \quad [2]$$

donde M es el número de meses analizados.

Finalmente, se procede a cuantificar las diferencias entre los buscadores a partir de la varianza de los datos. Para ello, se calcula la varianza aritmética entre los valores en bruto del tamaño web de cada mes en un determinado dominio web, para cada buscador.

b) Impacto (*visibility*)

En este caso se utilizan los buscadores *Yahoo!* y *Exalead*. Quedan excluidos *Google* y *Live* por no ofrecer en la actualidad los comandos necesarios para el cálculo de este parámetro.

Se propone el *factor de representatividad relativa media en visibilidad* (R_v), calculado a partir de la media normalizada de los valores de visibilidad de un determinado dominio académico web (v_{sn}) durante los 6 meses de estudio, de forma similar al R_s .

$$v_{sn} = \frac{x_{sn}}{N} \cdot 100 \quad [3]$$

$$R_v = \frac{\sum_{i=1}^M v_{sn}}{M} \quad [4]$$

En cuanto a la varianza en visibilidad entre buscadores, se calcula a partir de los datos en bruto mensuales de enlaces externos a cada portal.

Puesto que los dominios web mayores tienen estadísticamente más probabilidades de obtener un mayor número de enlaces externos, se procede a calcular el *WebIF* externo, de forma que se obtenga el número de enlaces externos por documentos indizados por el buscador.

El *WebIF*, introducido por Ingwersen (1998), se obtiene como la suma lógica del número de enlaces externos que apuntan a un determinado sitio web, dividido por el número de páginas de ese sitio en un momento determinado.

Pese a la existencia de ciertas críticas acerca del uso del este último indicador (Noruzi, 2006b), es la única forma actual de conocer la visibilidad por documento, siendo utilizado en diversos estudios enfocados a sedes académicas (Smith y Thelwall, 2004; Noruzi, 2006a).

Finalmente se muestra la curva de regresión y se calcula la correlación (Pearson y Spearman) entre R_s y R_v , tanto para *Exalead* como para *Yahoo!*, con el objeto de conocer la relación entre los dos indicadores propuestos, así como para verificar la influencia del tamaño de los dominios web en las medidas de visibilidad a lo largo del tiempo.

3. Resultados

3.1. Tamaño del portal

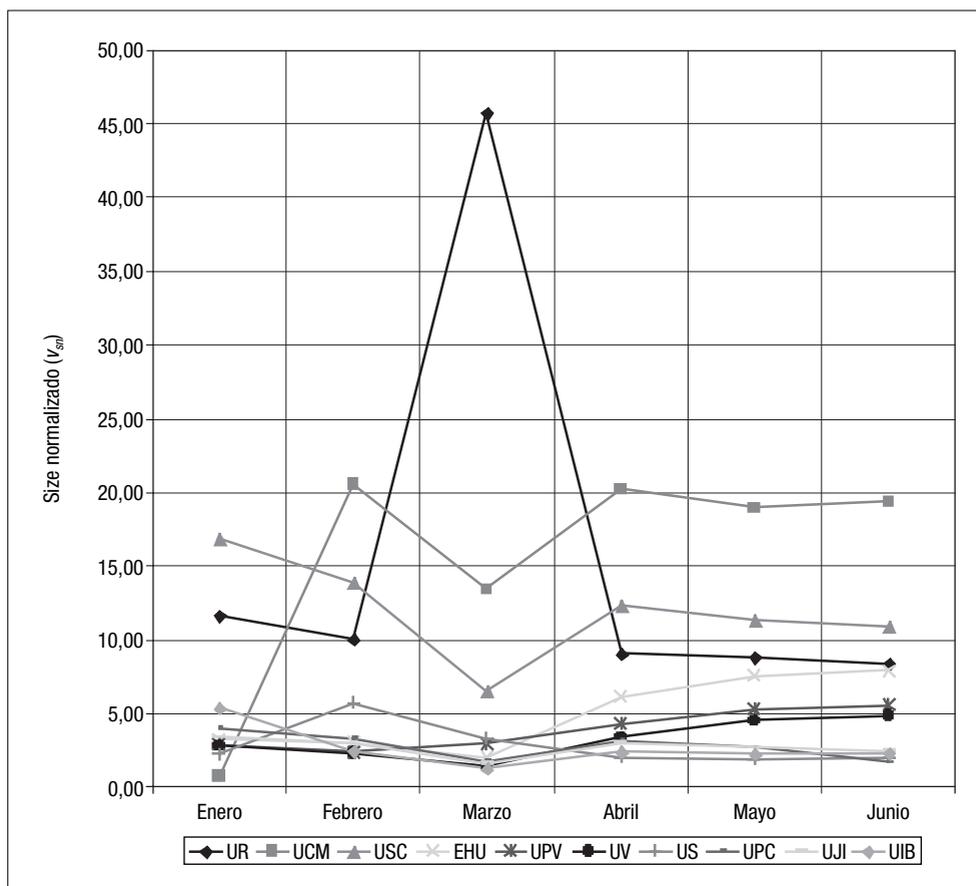
3.1.1. Google

La figura 1 muestra la evolución del tamaño normalizado v_{sn} correspondiente a los 10 dominios académicos con mayor R_s .

La UR es la universidad con mayor R_s en *Google*, debido fundamentalmente al pico registrado en marzo, que limita ligeramente la representatividad del resto de universidades. La amplitud del pico y los valores v_{sn} obtenidos en el resto de meses hace pensar en un comportamiento anómalo del buscador. La realización de varias catas por mes elimina la posibilidad de algún error en la toma de datos.

FIGURA 1

Evolución del tamaño normalizado v_{sn} en «Google»



Aparte de la UR, las universidades que presentan mejores resultados son la UCM (con un alto crecimiento en febrero que logra mantenerse a lo largo de los meses) y la USC, siendo estas 3 universidades las únicas en superar el valor de 10 en R_s . La EHU, por su parte, queda en cuarto lugar con un importante crecimiento desde abril, que la sitúa en valores cercanos a los de la UR.

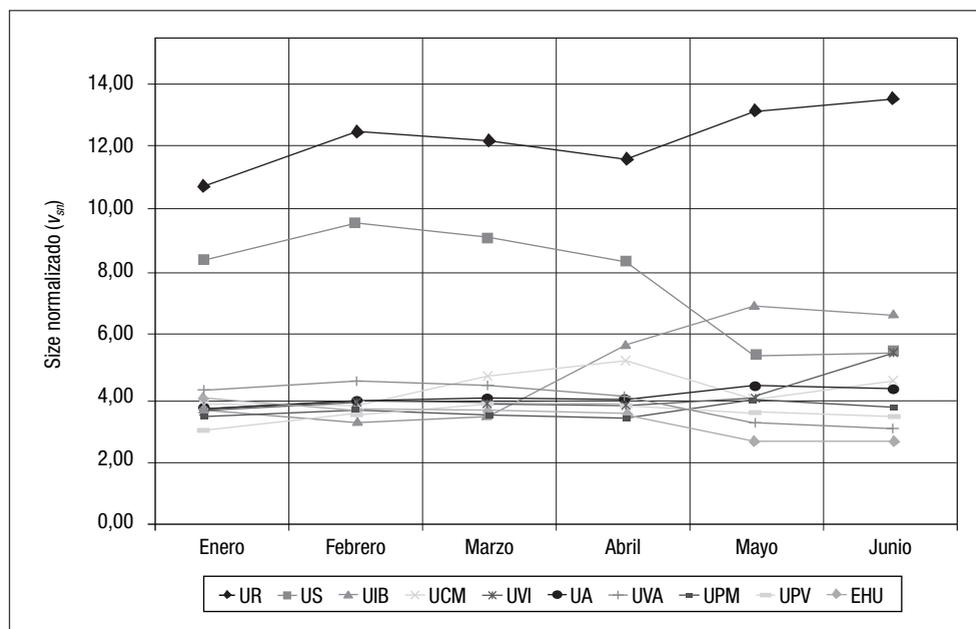
Finalmente, se destaca la presencia de 3 universidades públicas de la Comunidad valenciana (UPV, UJI, UV) entre las 10 universidades con mayor R_s , siendo la comunidad autónoma con más universidades en este rango de valores. Se detecta igualmente que los datos en bruto (no normalizados) de la UPV se repiten en enero-febrero y en marzo-mayo. Esto también ocurre en la UJI en marzo-abril. Aunque puede ser debido al proceso de cálculo de la media entre las diferentes catas del mes, resulta igualmente un comportamiento extraño por parte del buscador.

Asimismo, destaca la aparición de 2 universidades politécnicas (UPV y UPC) entre las diez primeras, así como la inclusión de la UIB en este grupo, por delante de otras instituciones académicas con mayor producción científica.

3.1.2. *Yahoo!*

Según *Yahoo!* (figura 2), la UR posee el mayor dominio web de todas las universidades españolas y, además, mantiene dicha posición durante todos los meses estudiados. La US es la segunda universidad en tamaño. Ambas universi-

FIGURA 2
Evolución del tamaño normalizado v_{sn} en «Yahoo!»



dades muestran una evolución muy similar de enero a abril; a partir de ese momento, sus evoluciones siguen trayectorias divergentes (crecimiento en la UR y fuerte caída en la US). Por otra parte, se detecta nuevamente una presencia notable de la UIB, con un fuerte crecimiento de marzo a mayo en sus valores v_{sn} .

Se muestra asimismo cómo 6 de las 10 primeras universidades aparecen también en el *top ten* mostrado para *Google* (figura 1), aunque en distinto orden. Las 4 universidades restantes son la UVI, UA, UVI y UPM, que logran en *Google* los puestos 11, 15, 14 y 16 respectivamente.

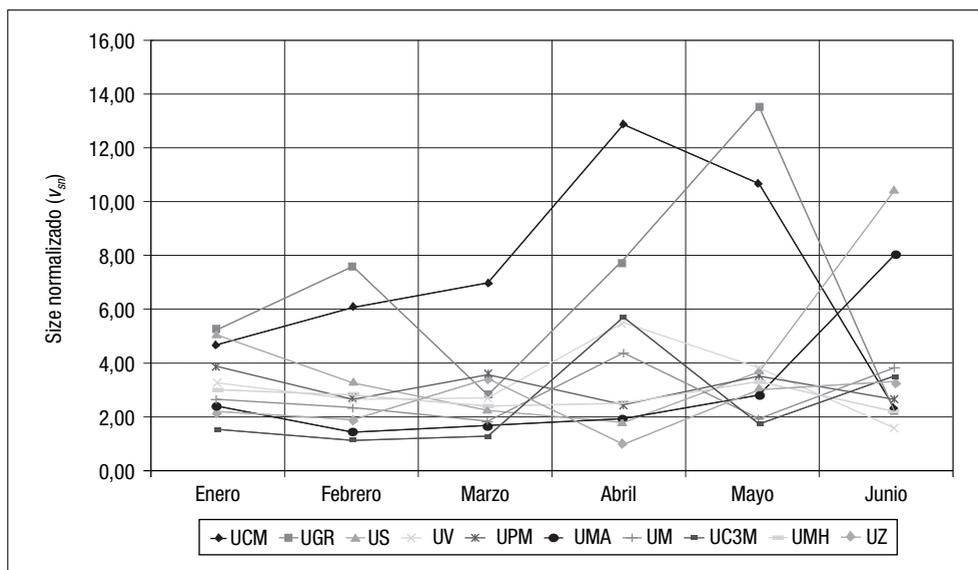
El valor más inesperado corresponde a la UV (tabla III), infrarrepresentada en *Yahoo!*, en la última posición. Los valores bajos se repiten además durante todos los meses, por lo que no se debe a un error puntual en la toma de datos, sino a la mala cobertura del buscador o a malas prácticas en el dominio web, aunque esto último también debería penalizar en el resto de buscadores, hecho que no ocurre.

Este hecho es detectado asimismo por Ortega y Aguillo (2007), quienes no encuentran información de la UV en *Yahoo! Search*. Se desconoce la causa de este problema pero supone una merma importante de la presencia de esta universidad en la web.

3.1.3. *Live*

Los datos referentes a *Live* (figura 3) muestran un comportamiento en la evolución del tamaño de los dominios web muy distinto al obtenido previamente

FIGURA 3
Evolución del tamaño normalizado v_{sn} en «Live»



para *Google* y *Yahoo!*, no sólo en las universidades con más v_{sn} y/o R_s , sino en los patrones de evolución, mostrando comportamientos menos lineales, con mayores picos y valles abruptos.

La UCM y la UGR son las universidades que acumulan más v_{sn} a lo largo de los meses (y por tanto mayor R_s), debido fundamentalmente a grandes crecimientos puntuales (en abril para la UCM —detectado también en *Yahoo!*— y en mayo para la UGR), con las consiguientes caídas en los meses posteriores.

Destaca igualmente el amplio crecimiento en representatividad de la US y UM en junio, que puede deberse a las grandes caídas de la UCM y la UGR en dicho mes. Por otra parte, se detecta un bajo rendimiento de la UR en este buscador, ocupando la posición 36 en el ranking R_s , frente al primer puesto logrado en el resto de buscadores (tabla III).

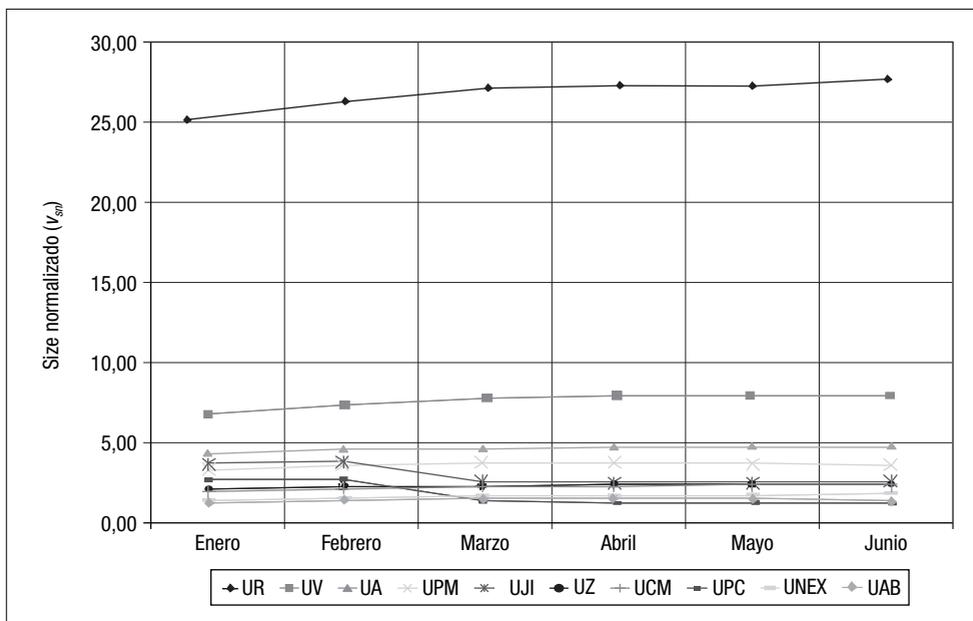
Asimismo, es significativa la procedencia geográfica de las universidades en este buscador, donde se cuenta con 3 universidades de la Comunidad de Madrid (UCM, UAM y UPM), otras 3 andaluzas (US, UGR y UM) y 2 valencianas (UV y UMH), así como la ausencia de universidades catalanas.

3.1.4. *Exalead*

La UR es de nuevo la universidad con el dominio web de mayor tamaño según *Exalead* (figura 4), situándose la UV en segunda posición pero bastante alejada.

FIGURA 4

Evolución del tamaño normalizado v_{sn} en «Exalead»



Los bajos resultados obtenidos en datos brutos por *Exalead* muestran una menor cobertura de este motor de búsqueda en los dominios académicos españoles. Asimismo, las tendencias de las universidades son muy estables, mostrando una significativa ausencia de grandes picos y valles en la evolución del tamaño web así como una ligera caída generalizada en todas las universidades. Sólo la UR (en junio), UA y UZ (en abril) y UCM y UNEX (en mayo), presentan valores más altos que los del mes anterior (para las 10 primeras universidades).

De nuevo, la procedencia geográfica de las 10 universidades con mayor *Rs* es digna de mención; en este caso se localizan 3 universidades de la Comunidad valenciana (UV, UA y UJI), 2 de la Comunidad de Madrid (UPM y UCM) y 2 catalanas (UPC y UAB).

3.1.5. Diferencias por buscador

a) *Rs por buscador*

En la tabla III se muestran los valores de *Rs* para todos los dominios web analizados en función de cada uno de los buscadores, y su correspondiente posición en dicho ranking.

TABLA III
Representatividad relativa media en tamaño

Universidades	Google		Yahoo!		Live		Exalead	
	Rs	Ranking	Rs	Ranking	Rs	Ranking	Rs	Ranking
EHU	5,00	4	3,53	10	2,52	11	1,66	18
UA	2,21	15	4,12	6	1,98	23	4,36	3
UAB	2,45	12	2,65	13	2,45	12	2,52	10
UAH	0,58	30	0,78	34	1,50	33	0,60	38
UAL	0,23	41	0,61	37	1,08	43	0,61	37
UAL	0,24	40	0,22	45	1,08	42	0,07	46
UAM	0,69	28	1,82	22	2,20	16	2,25	13
UB	1,06	23	1,17	28	2,00	22	1,65	19
UC	0,37	35	0,79	33	1,47	35	0,97	29
UC3M	1,20	21	2,32	17	2,72	8	0,74	33
UCA	0,66	29	1,24	27	1,31	38	0,70	34
UCLM	0,75	27	1,32	26	1,76	30	1,16	28
UCM	15,56	2	4,37	4	7,02	1	2,98	7
UCO	0,48	32	1,11	29	2,11	19	0,63	35

TABLA III (continuación)

Universidades	Google		Yahoo!		Live		Exalead	
	Rs	Ranking	Rs	Ranking	Rs	Ranking	Rs	Ranking
UDC	0,90	25	1,68	23	1,78	29	1,65	20
UDG	0,14	46	0,59	38	1,08	44	0,81	32
UDL	0,30	37	0,66	36	0,99	46	0,60	39
UGR	0,91	24	2,57	14	6,35	2	2,19	14
UHU	0,21	42	0,49	40	1,97	24	0,46	40
UIB	2,69	10	4,93	3	1,83	28	1,29	24
UJA	0,19	43	0,54	39	1,12	41	0,30	43
UJI	2,70	9	2,17	19	1,85	27	3,40	5
UL	0,18	44	0,45	41	1,50	34	0,44	42
ULL	0,32	36	0,83	31	1,67	32	0,88	30
ULPGC	1,85	17	1,67	24	1,33	37	0,85	31
UM	1,32	20	2,10	20	2,84	7	1,38	23
UMA	0,49	31	0,97	30	3,06	6	1,25	27
UMH	0,44	34	0,70	35	2,53	9	1,87	16
UNED	1,06	22	1,51	25	2,24	15	2,45	11
UNEX	0,44	33	2,00	21	2,08	21	2,61	9
UO	0,80	26	2,80	12	1,92	25	1,63	21
UPC	2,78	8	2,56	15	2,37	13	2,69	8
UPCT	0,14	45	0,24	43	1,01	45	0,20	44
UPM	1,97	16	3,72	8	3,10	5	3,78	4
UPNA	0,30	38	0,35	42	1,29	39	0,61	36
UPO	2,33	13	0,23	44	1,18	40	0,18	45
UPV	3,90	5	3,68	9	2,16	18	1,87	15
UR	15,63	1	12,23	1	1,40	36	26,58	1
URJC	0,27	39	0,79	32	2,27	14	0,46	41
US	2,89	7	7,68	2	4,34	3	2,39	12
USAL	1,45	19	2,55	16	2,18	17	1,69	17
USC	11,96	3	2,81	11	1,67	31	1,27	25
UV	3,23	6	0,00	46	3,22	4	7,55	2
UVA	2,33	14	4,03	7	1,85	26	1,47	22
UVI	2,64	11	4,18	5	2,10	20	1,27	26
UZ	1,75	18	2,24	18	2,52	10	3,01	6

Fuente: propia.

Pese a existir una gran diferencia entre los datos obtenidos para cada buscador, se observa una preponderancia de la UR (primera en *Google*, *Yahoo!* y *Exalead*) y de la UCM (primera en *Live* y segunda en *Google*). Los malos resultados de la UR en *Live* y de la UCM en *Exalead* parecen corresponder con la cobertura de estos buscadores.

No obstante, cada buscador ofrece algún resultado inesperado; *Google* ofrece valores excesivamente bajos para la UPM y altos para la USC y UPO; *Yahoo!* presenta valores muy altos para la UIB y bajos para la UV.

Live presenta valores relativamente altos para la UC3M, UGR, UHU, UM, UMA, UMH y URJC y bajos para la UA, UJI, UR. Finalmente, *Exalead* ofrece valores altos para la UJI, UNEX, UV y UZ y bajos para la UC3M.

b) *Varianza en tamaño entre buscadores*

Las diferencias mensuales entre los datos brutos pueden significar tendencias alcistas o decrecientes, pero cuando la varianza de datos es demasiado grande, los resultados pueden verse afectados por cambios en la cobertura de los buscadores, debidos a problemas técnicos o políticos, ajenos a las universidades. En ese sentido, *Google* presenta una variabilidad de datos muy grande, que hace plantear dudas acerca de la validez del comando, sobre todo en los casos de la UR y UCM.

Además, la diferencia en los resultados brutos entre los buscadores es muy elevada, lo que afecta a la varianza (a mayor cobertura, la varianza entre los datos tiene más probabilidades de ser más elevada). La tabla IV muestra la varianza del tamaño de los dominios para los 6 meses estudiados.

TABLA IV
Varianza de datos en tamaño

Universidades	Varianza (\$)			
	<i>Google</i>	<i>Yahoo!</i>	<i>Live</i>	<i>Exalead</i>
EHU	523.114,52	80.717,81	21.995,91	784,98
UA	150.938,91	16.942,06	20.139,80	937,40
UAB	284.705,94	15.106,15	14.162,71	660,84
UAH	14.158,82	1.400,36	8.584,56	234,22
UAL	8.873,26	4.400,62	5.489,69	93,46
UAM	6.806,86	1.931,92	23.662,89	952,15
UB	49.844,48	73.812,46	20.827,58	11.648,40
UBU	8.182,59	2.324,53	11.860,53	31,24
UC	7.231,05	5.023,47	7.859,53	45,70
UC3M	87.591,38	22.023,88	20.721,97	1.171,85
UCA	21.247,88	19.367,17	7.653,27	445,94
UCLM	13.301,00	12.220,67	12.654,83	290,45

TABLA IV (continuación)

Universidades	Varianza (S)			
	Google	Yahoo!	Live	Exalead
UCM	1.923.983,86	79.680,17	86.225,55	306,62
UCO	7.781,75	5.084,31	13.263,32	141,89
UDC	81.284,99	6.766,85	15.052,91	281,46
UDG	4.162,03	2.728,13	6.768,16	527,66
UDL	9.357,53	3.691,45	3.693,54	439,92
UGR	17.432,41	12.909,20	120.367,36	105,23
UHU	1.552,06	4.243,38	17.028,74	44,07
UIB	228.131,11	191.462,50	17.956,65	930,68
UJA	7.760,75	15.005,88	7.450,65	43,26
UJI	49.002,55	21.316,14	16.938,09	4.700,55
UL	4.531,89	3.472,02	16.178,96	86,00
ULL	5.347,48	3.041,85	17.189,01	179,08
ULPGC	148.794,79	9.076,37	8.234,36	401,70
UM	194.654,75	31.047,80	22.983,09	604,33
UMA	7.624,89	2.134,85	62.999,99	400,11
UMH	72.665,92	6.055,45	32.208,69	200,47
UNED	13.609,64	63.347,04	28.618,22	454,13
UNEX	5.557,78	34.964,82	19.142,75	318,92
UO	169.333,01	38.995,29	18.632,11	1.684,21
UPC	153.202,93	22.304,48	26.221,66	4.957,78
UPCT	7.604,09	1.314,34	2.132,62	22,15
UPM	40.226,10	5.967,42	24.509,36	902,26
UPNA	7.262,10	958,55	4.348,18	92,05
UPO	203.148,07	2.571,54	10.759,92	88,89
UPV	328.294,08	43.173,94	21.803,47	714,81
UR	6.763.679,14	75.487,97	5.246,48	3.444,71
URJC	5.415,92	3.132,38	32.140,44	237,49
US	474.475,03	286.389,83	85.456,92	1.919,72
USAL	43.071,32	9.372,18	16.993,80	1.809,57
USC	400.017,36	19.699,94	11.361,11	239,98
UV	259.900,25	10,91	27.716,95	632,67
UVA	168.641,30	92.965,92	7.653,90	268,00
UVI	253.766,73	68.784,42	24.181,43	142,96
UZ	189.585,51	8.269,66	23.508,27	783,52
Promedio	291.888,17	31.102,13	22.403,91	987,03

Se observa cómo los datos de *Google* son los que presentan, en términos generales, un mayor valor para todas las universidades mientras que *Exalead* presenta los valores más bajos (lo que corresponde con sus coberturas). *Yahoo!* y *Live* alternan sus valores en función de la universidad, aunque el primero presenta en promedio una mayor varianza que el segundo.

No obstante, no sólo es importante la varianza de datos (pues se ve influida por el tamaño de los dominios y cobertura de los buscadores), sino las diferentes tendencias en la evolución de los resultados mostradas en las figuras 1 a 4.

3.2. Visibilidad

3.2.1. *Yahoo!*

El cálculo del número de enlaces externos entrantes muestra que la UR es la universidad que mayor visibilidad alcanza, con niveles muy superiores al resto (tabla V). La UCM y la UA se sitúan destacadas en segundo y tercer puesto respectivamente.

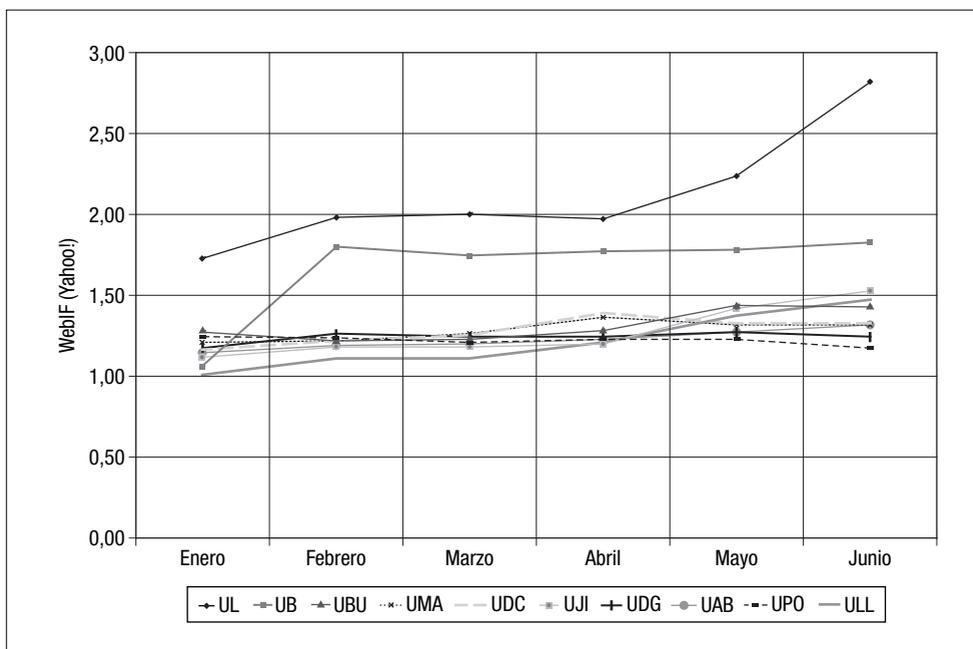
El ranking obtenido concuerda *grosso modo* con el correspondiente en tamaño para este buscador, aunque con algunas diferencias, como la inclusión de la UV (a pesar de no detectar páginas, este buscador sí detecta enlaces a dichas páginas), la UAB y la UGR. Este hecho muestra un principio de relación entre el tamaño y la visibilidad web.

TABLA V
Visibilidad (Yahoo!)

Universidad	Enero	Febrero	Marzo	Abril	Mayo	Junio	Media
UR	1.055.445	1.232.151	1.258.476	1.324.016	1.333.421	1.456.372	1.276.646,83
UCM	654.327	637.646	776.709	848.495	723.443	876.343	752.827,17
UA	646.139	666.416	691.960	727.483	728.590	750.700	701.881,33
UPV	447.448	471.042	570.842	571.304	582.892	606.002	541.588,33
UV	540.077	526.067	541.746	541.148	540.284	543.878	538.866,67
UVI	505.306	518.865	509.585	517.653	397.883	629.057	513.058,17
US	482.836	497.970	505.027	520.839	522.654	540.805	511.688,50
UPM	463.697	462.309	481.463	513.040	508.016	535.608	494.022,17
UAB	477.595	468.612	480.158	502.775	475.722	501.918	484.463,33
UGR	406.251	412.283	425.163	418.005	421.850	441.222	420.795,67

El cálculo del *WebIF* para *Yahoo!* se muestra en la figura 5, donde se obtienen los resultados para las 10 primeras universidades a excepción de la UV, que ha sido eliminada debido a los problemas detectados en tamaño, que propician un crecimiento artificial en su *WebIF*.

FIGURA 5
WebIF en Yahoo!



Descontando la UV, 23 universidades han logrado un *WebIF* superior a 1 y el resto (22), inferior a la unidad. Este hecho pone de manifiesto que prácticamente el 50% de los dominios web españoles ni siquiera llega a un ratio de un enlace por página en *Yahoo!*, lo que muestra la gran cantidad de páginas que no reciben enlaces y, por tanto, no tienen impacto en la Web, al menos en este importante buscador. La UL y la UB son las universidades con un mayor *WebIF* según *Yahoo!* en el período estudiado, seguidas de la UBU, hecho que sorprende por no corresponder a universidades que destaquen en tamaño y/o visibilidad en dicho buscador.

El resto de universidades se mantiene en un margen de valores entre el 1 y 1,5, con una tendencia ligeramente alcista en los últimos meses, sobre todo en la UJI y ULL. Asimismo, el crecimiento en impacto web de la UL es muy intenso a partir del mes de marzo.

3.2.2. *Exalead*

Los valores de visibilidad en *Exalead* se detallan en la tabla VI. Se obtienen cifras de enlaces mucho más reducidas que en *Yahoo!*, lo cual es comprensible debido a las diferencias que presentan ambos en cobertura de la Web.

TABLA VI
Visibilidad (Exalead)

Universidad	Enero	Febrero	Marzo	Abril	Mayo	Junio	Media
UGR	167.330	180.255	183.702	186.591	186.538	272.448	196.144,00
UCM	58.375	57.399	55.287	54.894	55.948	59.508	56.901,83
UV	51.392	50.765	47.672	47.571	47.293	46.476	48.528,17
UR	15.871	13.304	9.211	11.170	12.870	260.768	53.865,67
UPV	48.041	44.332	40.491	37.379	36.615	29.746	39.434,00
UPC	39.383	38.541	37.458	37.143	37.097	37.504	37.854,33
UAB	36.440	36.372	34.647	34.130	33.246	39.124	35.659,83
UPM	31.807	31.374	31.664	32.646	32.499	40.378	33.394,67
UA	30.697	30.860	30.210	30.285	29.325	44.851	32.704,67
US	30.726	30.448	29.262	29.156	29.037	34.217	30.474,33

La relación de la tabla VI con la tabla V (visibilidad en *Yahoo!*) es relativamente alta, pues 9 de las 10 universidades con más visibilidad en *Yahoo!* están igualmente entre las 10 con más enlaces externos en *Exalead*, aunque en posiciones distintas. Estos datos consolidan un *cluster* de universidades más visibles en los buscadores.

La relación con los datos de la figura 4 (tamaño en *Exalead*) también es relativamente alta (7 de 10 universidades aparecen en ambos rankings), aunque la UGR presenta una clara excepción.

En este caso, la UGR es la universidad con mayor número de enlaces, seguida de la UCM y de la UV. De nuevo destaca la alta presencia de universidades andaluzas (US, UGR), valencianas (UV, UPV y UA), madrileñas (UCM y UPM) y también catalanas (UPC y UAB). Por otra parte, el número de enlaces de la UR no guarda correspondencia con la presencia en este buscador (UR es la universidad con mayor *Rs* en *Exalead*).

En lo que respecta al impacto web, la figura 6 detalla la evolución de este indicador a lo largo de los meses estudiados.

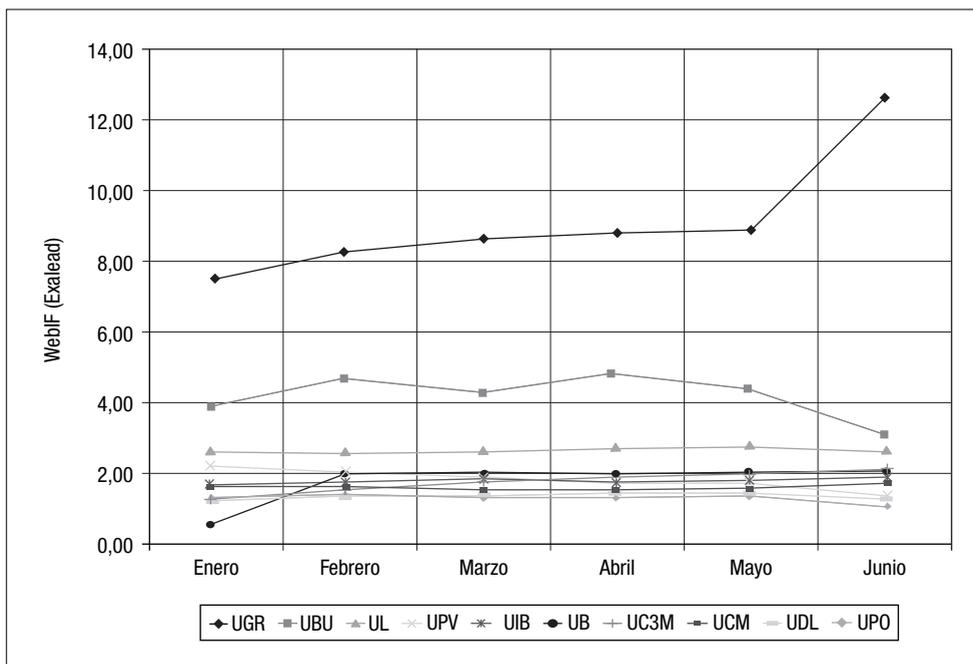
En este caso, y a diferencia de *Yahoo!*, la universidad con mayor visibilidad también logra los valores máximos de *WebIF*. Así, la UGR ejerce un claro predominio en este parámetro, pese a detectar algún valor de pico exagerado, como el localizado en junio debido al incremento de enlaces obtenidos ese mes.

A pesar de ello, el *WebIF* permanece muy estable en todas las universidades (generalmente en torno a un valor de 2), excepto algunos cambios muy puntuales, como el crecimiento en la UB en febrero o las caídas de la UB, UPV y UPO en junio.

Otras universidades como la UBU, UL y UPO, logran tanto en *Yahoo!* como en *Exalead* unos factores de impacto relativamente altos, debido a los bajos valores logrados en tamaño web en relación al número de enlaces obtenidos.

FIGURA 6

WebIF en Exalead



3.2.3. Diferencias en visibilidad por buscador

a) *Rv* por buscador

En la tabla VII se muestra el *Rv* para los dominios académicos web, así como el ranking global de cada universidad en cada buscador.

Los resultados muestran una buena correspondencia entre los rankings ocupados por las universidades en ambos buscadores, aunque existen algunas excepciones, como la UGR, UMA, USAL, UVA, UVI, UB, UDG y UV.

De esta forma, las diferencias mostradas permiten conocer el bajo rendimiento de una universidad en un determinado buscador, mientras que las similitudes en las posiciones ocupadas por las universidades permite la confección de clusters o agrupaciones más refinadas. En ese sentido, las universidades que consiguen figurar dentro del *top ten* en ambos buscadores son la UCM, UGR, UR, US, UPM, UPV, UAB y UA.

TABLA VII
Representatividad relativa media en visibilidad (Rv)

Universidades	<i>Yaboo!</i>		<i>Exalead</i>	
	<i>Rv</i>	Ranking	<i>Rv</i>	Ranking
UAM	2,21	20	2,20	16
UC3M	1,58	28	1,44	24
UCM	5,57	2	5,45	2
UAH	0,81	38	0,80	36
UAL	0,61	39	0,42	40
UBU	0,32	44	0,25	45
UCA	1,00	33	0,64	37
UC	0,96	35	0,93	33
UCLM	1,62	27	1,16	30
UCO	1,40	29	0,90	34
UNEX	1,71	26	0,88	35
UGR	3,12	10	18,53	1
UHU	0,55	41	0,42	41
UJA	0,51	42	0,33	43
ULL	1,11	31	1,25	29
UR	9,45	1	4,26	4
ULPGC	1,76	24	1,08	31
UL	1,04	32	1,27	28
UMA	1,38	30	1,78	20
UM	2,23	19	1,99	18
UO	1,91	23	2,15	17
USAL	3,12	11	1,77	21
US	3,80	7	2,91	10
UVA	4,00	4	1,72	22
UZ	2,66	13	2,82	12
EHU	2,60	14	2,29	15
UMH	0,98	34	0,63	38
UNED	1,74	25	1,71	23
UPO	0,32	45	0,29	44
UPCT	0,25	46	0,14	46
UPM	3,66	8	3,18	8
UPV	4,01	5	3,82	5

TABLA VII (continuación)

Universidades	<i>Yahoo!</i>		<i>Exalead</i>	
	<i>Rv</i>	Ranking	<i>Rv</i>	Ranking
UPNA	0,39	43	0,33	42
URJC	0,86	36	0,60	39
UDC	2,37	17	1,43	25
USC	2,52	15	1,81	19
UVI	3,80	6	1,31	26
UAB	3,60	9	3,41	7
UA	5,20	3	3,09	9
UB	2,03	22	2,82	11
UDG	0,81	37	1,28	27
UIB	2,40	16	2,57	14
UDL	0,59	40	0,98	32
UV	2,25	18	4,67	3
UJI	2,11	21	2,65	13
UPC	3,09	12	3,64	6

b) Varianza en visibilidad entre buscadores

En la tabla VIII se detallan los valores de varianza en los datos de visibilidad ofrecidos por *Yahoo!* y *Exalead*. Los valores son más elevados en *Yahoo!*, lo cual se corresponde con su mayor volumen de datos, al igual que ocurría con el tamaño web (tabla IV).

TABLA VIII
Varianza de datos en visibilidad

Universidades	Varianza (V)	
	<i>Yahoo!</i>	<i>Exalead</i>
UAM	11.348,53	1.490,55
UC3M	11.430,33	398,33
UCM	90.195,39	1.671,25
UAH	5.880,01	490,30
UAL	12.409,88	475,64
UBU	3.489,21	289,24
UCA	9.374,29	646,42
UC	6.644,65	314,47
UCLM	16.373,37	409,72

TABLA VIII (continuación)

Universidades	Varianza (V)	
	<i>Yaboo!</i>	<i>Exalead</i>
UCO	7.190,14	182,22
UNEX	7.921,61	373,61
UGR	11.024,37	34.745,13
UHU	2.525,75	562,44
UJA	1.752,75	292,69
ULL	16.630,16	669,13
UR	121.798,05	92.551,81
ULPGC	7.358,56	1.073,73
UL	17.526,56	188,96
UMA	9.511,36	1.027,05
UM	12.319,80	1.360,40
UO	4.351,09	1.279,76
USAL	29.168,70	1.859,71
US	18.782,90	1.795,61
UVA	5.858,24	1.400,40
UZ	7.942,69	1.033,76
EHU	18.926,66	2.082,78
UMH	8.989,97	211,03
UNED	3.354,90	382,08
UPO	2.557,82	77,80
UPCT	3.131,56	90,39
UPM	26.987,53	3.155,12
UPV	59.767,74	5.848,29
UPNA	1.898,98	265,08
URJC	8.890,66	144,02
UDC	25.329,53	707,09
USC	19.229,02	1.339,53
UVI	66.893,00	518,04
UAB	13.123,57	1.928,26
UA	37.003,36	5.453,86
UB	40.978,27	6.302,75
UDG	647,10	1.132,96
UIB	4.759,52	1.180,13
UDL	1.095,81	569,84
UV	5.579,68	1.852,53
UJI	5.086,52	5.706,87
UPC	8.040,98	833,44
Promedio	17.632,19	4.051,40

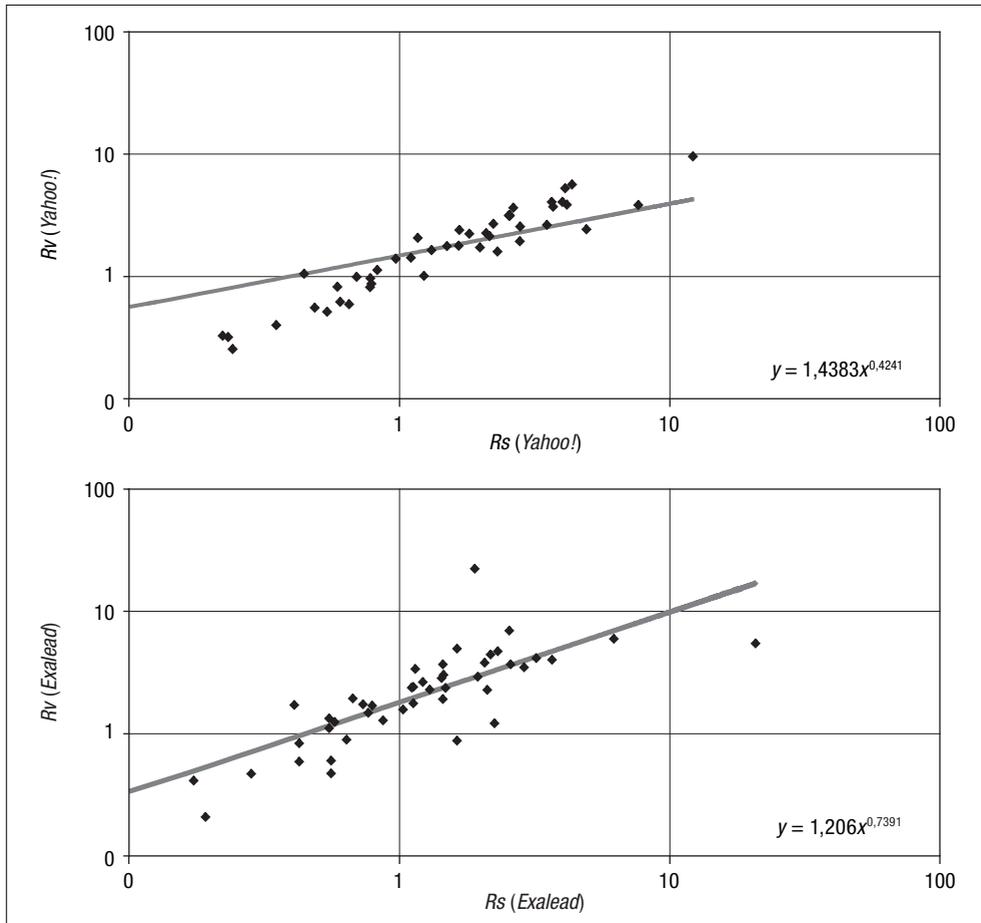
Estos resultados confirman la alta variabilidad de datos de la UR (tanto en *Yahoo!* como en *Exalead*), de la UCM (fundamentalmente en *Yahoo!*) y de la UGR (en *Exalead*).

No obstante, hay que tomar con precaución los datos de varianza, pues valores mínimos muestran una estabilidad en los datos del buscador, pero también una carencia de crecimiento en la Web por parte de dichas universidades.

3.3. Regresión y correlación entre *Rs* y *Rv*

En la figura 7 se muestra la regresión potencial obtenida para *Rs* y *Rv* (y por tanto, entre tamaño y visibilidad), en *Yahoo!* y *Exalead*. Se aprecia como a mayor

FIGURA 7
*Regresión potencial entre *Rs* y *Rv* en *Yahoo!* y *Exalead**



R_s aumenta el valor de R_v en prácticamente todos los dominios web académicos españoles, a lo largo del tiempo. Los ejes vertical (R_v) y horizontal (R_s) se muestran en escala logarítmica con el objetivo de evitar que los valores extremos determinen el ajuste de la regresión.

Por otra parte, el coeficiente de correlación de Pearson entre R_s y R_v es de 0,90 para *Yahoo!* y de 0,24 para *Exalead*. No obstante, en este último caso, los datos están contaminados por la UR y UGR; eliminando estos dominios del cálculo, el coeficiente de correlación de Pearson es de 0,77.

Puesto que la distribución de frecuencias de tamaño y visibilidad por dominio no sigue una distribución normal, sino potencial, se procede a utilizar el coeficiente de correlación de Spearman, más adecuado en este tipo de distribución. De esta forma, *Yahoo!* logra un valor muy alto ($r = 0,9973$), al igual que *Exalead* ($r = 0,9507$). Si eliminamos los valores de la UR y UGR, como se ha hecho anteriormente, *Exalead* logra aumentar su coeficiente ($r = 0,9976$).

La figura 7 permite además observar de forma clara los comportamientos de cada dominio web, detectando rendimientos que quedan fuera de lo normal. En ese sentido, la correlación en *Yahoo!* muestra de forma gráfica los problemas en la detección de tamaño en la UV (coordenadas [0,00; 2,25]), los altos valores de la UR [12,23; 9,45], tanto en visibilidad como en tamaño, el alto rendimiento de la UA [4,12; 5,20] y UCM [4,37; 5,57], y la poca visibilidad de la UIB [4,93; 2,40] y la US [7,68; 3,80].

Asimismo, los datos relativos a *Exalead* muestran la poca visibilidad de la UR en función del tamaño detectado [26,58; 4,26], así como el caso contrario para la UGR [2,19; 18,53]. Otras universidades que destacan son la UCM [2,98; 5,45] y la UV [7,55; 4,67], quienes alcanzan una gran visibilidad en relación a su tamaño web.

4. Discusión

La correlación detectada entre R_s y R_v implica una clara relación entre la visibilidad y tamaño de los dominios. Esto es igualmente detectado por Thelwall y Aguillo (2003), quienes demuestran que las grandes sedes tienden a atraer más enlaces.

En base a esto, los autores afirman que «si el modelo de interconexión esta relacionado con la productividad investigadora, como en el Reino Unido, entonces la desviación de una tendencia lineal puede representar instituciones relativamente productivas o improproductivas».

Es decir, una universidad con una visibilidad mayor a la que le correspondiera por tamaño, se relacionaría con una producción científica mayor que la media, pero esta interconexión con la actividad científica no es un fenómeno universal. De hecho, Thelwall y Aguillo discuten dos situaciones posibles:

- Que exista la interconexión y por tanto la productividad científica de la institución sea mayor o menor según su tamaño y visibilidad.

- Que grandes colecciones puntuales de páginas puedan estar inflando artificialmente el tamaño de las sedes, pero no se correspondan con cambios en la productividad científica.

Precisamente, este último punto corresponde con los resultados obtenidos en este trabajo para la UR (a través de la indización del servicio *Dialnet*¹⁵) y de la UCM (debido a *Compludoc*¹⁶). Este mismo fenómeno se ha detectado igualmente en la presencia de estas universidades en *Google Scholar* (Orduña-Malea y otros, 2009). La falta de comandos en *Google* para estudiar la visibilidad impide un estudio directo entre ésta y la presencia en *Google Scholar*

La alta correlación encontrada entre el tamaño y la visibilidad de los dominios web afecta igualmente al cálculo del *WebIF*. Así, las universidades con mayor tamaño también logran proporcionalmente visibilidades más altas, por lo que este indicador (obtenido a partir de la división entre visibilidad y tamaño), se mantiene constante, favoreciendo a universidades con mucha visibilidad y poco tamaño.

Estos resultados concuerdan con los obtenidos por Noruzi (2006a y 2006b) así como por otros autores, acerca de la poca información que este indicador aporta, pese a que constituye la única posibilidad de conocer, al menos de forma aproximada, los enlaces por documento.

Finalmente, resulta necesario comentar que tan sólo se ha utilizado un dominio por universidad, tal como muestra en la tabla I. Algunas universidades, como las catalanas, usan diversos dominios académicos (por ejemplo, .es y .cat). Este aspecto, ya debatido y criticado en la bibliografía (Aguillo, 2008), puede representar una limitación técnica a la hora de tomar en cuenta los resultados obtenidos, por lo que deberá ser tenido en cuenta. No obstante, la decisión de tomar en consideración o no dichos dominios alternativos queda lejos de estar solucionada, pues generalmente éstos presentan en su mayoría documentación repetida o duplicada, hecho que está provocando que en las últimas ediciones del RWWU sólo se tenga en cuenta el dominio principal de cada universidad. Sería necesario el desarrollo de una aplicación capaz de automatizar la detección de contenido duplicado entre documentos alojados en dominios alternativos para perfeccionar la metodología y solucionar así este problema.

5. Conclusiones

Las diferencias en tamaño entre las sedes académicas web son muy grandes según el buscador utilizado, por lo que el uso de un único buscador no permite conocer de forma fiable el tamaño real de un dominio web. Además, el empleo de valores combinados a partir de la media obtenida en cada buscador no ofrece resultados fiables, dada la varianza de datos obtenida entre buscadores así como los diferentes índices obtenidos de *Rs*.

Por otra parte, en términos cuantitativos, *Google* es el motor que ofrece una mayor cantidad de resultados, seguido de *Yahoo!*, *Live* y *Exalead*. Este orden

coincide en la varianza de los datos obtenidos, lo cual es lógico, aunque se advierten ciertos comportamientos anómalos en ciertas universidades.

Se observa una disminución en tamaño de prácticamente todos los dominios en *Exalead*, lo que podría significar que las universidades españolas están perdiendo representatividad en este buscador y/o que la cobertura de éste disminuye paulatinamente. En cualquier caso, dados sus resultados, se plantea la idoneidad o no del uso de este buscador en estudios nacionales.

No obstante, los cambios en la evolución del tamaño de los dominios web pueden ser debidos tanto a la cobertura y funcionamiento de los buscadores como a acciones específicas llevadas a cabo por las universidades. La discriminación de estos fenómenos no se puede realizar mediante un simple análisis de los buscadores, por lo que las razones que pueden explicar subidas o bajadas son difíciles de contextualizar.

Además, la falta de gestión de la documentación web docente y su mayor obsolescencia podrían ser las causas de ciertos cambios en las evoluciones de tamaño detectadas en los dominios web. La carencia de comandos adecuados en los motores de búsqueda para discriminar información científica, docente y administrativa dentro de los dominios web impide conocer la influencia de estos distintos tipos de documentación, tanto en el tamaño y visibilidad de los dominios, como en su perdurabilidad. La existencia de sistemas como *Google Scholar* ayuda en relación a la documentación científica, pero no así con el resto.

Respecto a la visibilidad de los buscadores, se observan datos más constantes y estables en cada buscador, con varianzas menores que las detectadas en tamaño, aunque igualmente significativas, tanto en magnitud como en el orden de los dominios, por lo que también es desaconsejable el uso de un único buscador para conocer el verdadero impacto de las universidades.

La indización por parte de los buscadores de servicios como *Dialnet* en la UR y *Compludoc* en la UCM provoca la obtención de resultados artificiales tanto en tamaño como en visibilidad, pues la documentación alojada en estos servicios no es producida por dicha universidad, es decir, no es creada por autores pertenecientes a dichas instituciones y por tanto no se correlacionan con la producción científica de dichas universidades, como sí lo hacen los repositorios institucionales.

Este hecho afecta en la representatividad del resto de dominios, como demuestran los datos obtenidos. Sin embargo, no todos los buscadores indizan igual estos servicios, por lo que las políticas acerca del tratamiento de éstos por parte de los buscadores deberían normalizarse.

El *WebIF*, por su parte, ha proporcionado resultados diferentes en *Yahoo!* y *Exalead*. En el primero, se han constatado diferencias respecto a los resultados de visibilidad, mientras que en el segundo, las posiciones de las primeras universidades coinciden más. Pese a todo, la utilización de este indicador ha favorecido a universidades sin mucho tamaño pero con una visibilidad media, dada la alta correlación entre tamaño y visibilidad, lo que confirma la poca idoneidad de su uso.

Pese a que en los estudios evolutivos de la Web se han detectado dependencias en la estabilidad de las páginas web de variables como dominio, edad, contenidos, etc., este trabajo sólo pretende medir la evolución del tamaño y visibilidad de los dominios académicos web de las universidades públicas españolas y proponer una serie de indicadores (R_s y R_v) que pueden mejorar la construcción e interpretación de rankings web de universidades.

Aun así, se estima conveniente la realización de futuros estudios acerca de la influencia de las variables de perdurabilidad en el contexto académico español, como por ejemplo un análisis de la persistencia del dominio *.es* (*top level domain* empleado mayoritariamente), la influencia de repositorios institucionales, antigüedad de los dominios web, cantidad de documentación no científica, etc., temas todos ellos fuera del propósito de este trabajo pero que pueden ayudar a comprender e contextualizar los resultados obtenidos.

Finalmente, dado que el objeto de este estudio son los dominios web de instituciones académicas públicas, y por tanto financiadas con dinero público, se abre la duda acerca de si la información exacta de su tamaño real (Internet pública) debiera ser ofrecida por la propia universidad. Por ello, se estima conveniente la creación de una línea de trabajo dirigida al diseño y desarrollo de algún tipo de metadato, generado por la propia universidad, que pudiera ser leído por los distintos buscadores y que ofreciese esta información de forma actualizada.

6. Notas

1. El *U.S News & World Report* se comienza a publicar en 1983, pero no es hasta 1987 cuando se convierte en anual. <http://www.usnews.com/sections/rankings/index.html>.
2. *Leiden Ranking*. <http://www.cwts.nl/ranking/LeidenRankingWebSite.html> [consulta 20-10-2009].
3. Las categorías de indicadores propuestas son las siguientes: *Beginning characteristics; Learning inputs (resources and staff); Learning outputs; Final outcomes; Reputation; Research*.
4. Se trata del *Academic Ranking of World Universities (ARWU)*, desarrollado en la Shanghai Jiao Tong University. <http://www.arwu.org/> [consulta 20-10-2009].
5. *International Observatory on Academic Rankings and Excellence (IREG)*. <http://www.ireg-observatory.org/> [consulta 20-10-2009].
6. *Ranking Web of World Universities*. <http://www.webometrics.info/> [consulta 20-10-2009].
7. *Web Popularity Ranking*. <http://www.4icu.org/top200/> [consulta 20-10-2009].
8. *Ranking Universitario de transparencia Web*. <http://www.universidad.edu.co> [consulta 20-10-2009].
9. *Ranking Colleges using Google and OSS*. <http://vcmike.blogspot.com/2006/01/ranking-colleges-using-google-and-oss.html> [consulta 20-10-2009].

10. <http://www.universitymetrics.com/g-factor> [consulta 20-10-2009].
11. http://www.webometrics.info/best_practices.html [consulta 20-10-2009].
12. *Conferencia de Rectores de las Universidades Españolas (CRUE)*. <http://www.crue.org/> [consulta 20-10-2009].
13. Se obtienen datos de las 49 universidades públicas españolas, a excepción de la *Universidad Pompeu Fabra*, *Universidad Rovira i Virgili* y *Universidad Internacional de Andalucía*, de las que no se obtienen datos debido a problemas técnicos, por lo que no se consideran en ningún momento.
14. A principios de junio se pone en marcha, por parte de *Microsoft*, el buscador *Bing*, con el que sustituye a *Live*. Por ello, la recuperación de datos de enero a mayo se realiza en *Live*, mientras que los datos de junio se recogen con *Bing*.
15. *Dialnet*. <http://dialnet.unirioja.es/> [consulta 20-10-2009].
16. *Compludoc*. <http://europa.sim.ucm.es/compludoc/> [consulta 20-10-2009].

7. Bibliografía

- Aguillo, I. F., y Granadino, B. (2006): Indicadores web para medir la presencia de las universidades en la red. *Revista de universidad y sociedad del conocimiento*, vol. 3 (1), 68-75.
- Aguillo, I. F.; Granadino, B.; Ortega, J. L, y Prieto, J. A. (2006): Scientific research activity and Communications measured with cybermetrics indicators. *Journal of the American Society for information science and technology*, vol. 57 (10), 1296-1302.
- Aguillo, I. F.; Ortega, J. L., y Fernández, M. (2008): Webometric Ranking of World Universities: introduction, methodology, and future developments. *Higher education in Europe*, vol. 33 (2/3), 233-244.
- Altbach, P.G. (2006): GATS Redux: The WTO and Higher Education Returns to Center Stage. En Altbach, Philip G. (ed.), *International Higher education: reflexions on Policy and Practice*. Center for International Higher Education, Boston, EE.UU.
- Bar-Ilan, J. (1999), Search engine results over time – A case study on search engine stability, *Cybermetrics*, 2/3. Available: <http://www.cindoc.csic.es/cybermetrics/articles/v2i1p1.html>.
- Bar-Ilan, J. (2002): Methods for measuring search engine performance over time, *Journal of the American Society for Information Science and Technology*, 53 (4): 308-319.
- Bar-Ilan, J. (2003): The use of web search engines in information science research. *Annual review of information science and technology*, vol. 38, 231-288.
- Bar-Ilan, J., y Peritz, B. C. (2004): Evolution, continuity, and disappearance of documents on a specific topic on the Web: A longitudinal study of informetrics. *Jasist*, vol. 55(11), 980-990.
- Berlin Principles on Ranking of Higher Education Institutions* (2006): Institute for Higher Education Institut (IHEP). <http://www.ihep.org/assets/files/publications/a-f/BerlinPrinciplesRanking.pdf> [consulta 20-10-2009].

- Brewington, B. E., y Cybenko, G. (2000): How dynamic is the Web? *Computer Networks*, vol. 33 (1-6), 257-276.
- Cho, Y., y García-Molina, H. (2000): The evolution of the web and implications for an incremental crawler. *Proceedings of the 26th International Conference on Very Large Data Bases*, 200-209.
- Fetterly, D.; Manasse, M.; Najork, M., y Wiener, J. (2003): A large scale study of the evolution of web pages, *Proceedings of the Twelfth International Conference on World Wide Web*, 669-678.
- Hazelkorn, E. (2007): The impact of league tables and ranking systems on Higher Education decision making. *Higher education management and policy*, vol. 19 (2).
- Ingwersen, P. (1998): The calculation of web impact factors. *Journal of Documentation*, vol. 54(2), 236-243.
- Koehler, W. (1999): An analysis of web page and web site constancy and permanence, *Journal of the American Society for Information Science*, vol. 50 (2), 162-180.
- Koehler, W. (2002): Web page change and persistence -4- year longitudinal web study, *Journal of the American Society for Information Science and Technology*, vol. 53 (2), 162-171.
- Koehler, W. (2004): A longitudinal study of Web pages continued a consideration of document persistence, *Information Research*, vol. 9 (2). <http://informationr.net/ir/9-2/paper174.html> [consulta 20-10-2009].
- Lawrence, S., y Giles L. (1999): Accesibility of information on the Web. *Nature*, vol. 400, 107-109.
- Noruzi, A. (2006a): The Web Impact Factor: a critical review. *The Electronic Library*, vol. 24 (4), 490-500.
- Noruzi, A. (2006b): Web presence and impact factors for Middle-Eastern countries. *Online*, vol. 30 (2), 22-28.
- Orduña-Malea, E.; Serrano-Cobos, J., y Lloret-Romero, N. (2009): Las universidades públicas españolas en Google Scholar: presencia y evolución de su publicación académica web. *El profesional de la información*, vol. 18 (5), 493-500.
- Ortega, J. L., y Aguillo, Isidro-F. (2007): La web académica española en el contexto del Espacio Europeo de Educación Superior: estudio exploratorio. *El profesional de la información*, vol. 16 (5), 417-425.
- Ortega, J. L.; Aguillo, I. F., y Prieto, J. A. (2006): Longitudinal study of contents and elements in the scientific Web environment. *Journal of Information Science*, vol. 32 (4), 344-351.
- Payne, N., y Thelwall, M. (2007): A longitudinal study of academic webs: growth and stabilization. *Scientometrics*, vol. 71 (3), 523-539.
- Rocki, M. (2005): Statistical and mathematical aspects of rankings: lessons from Poland. *Higher education in Europe*, v. 30 (2), 173-181.
- Rousseau, R. (1999), Time evolution of the number of hits in keyword searches on the Internet, *Post Conference Seminar – Cybermetrics'99 at the Seventh International Conference on Scientometrics and Informetrics*, July 9 1999, Colima, Mexico. Available: <http://www.cindoc.csic.es/cybermetrics/cybermetrics99.html> [consulta 20-10-2009].
- Smith, A., y Thelwall, M. (2004): Web impact factor for Australasian universities. *Scientometrics*, vol. 54 (3), 363-380.

- Stuart, D. (1995): Reputational rankings: background and development. *New directions for institutional research*, vol. 88.
- Thelwall, M. (2001): Extracting macroscopic information from web links, *Journal of the American Society for Information Science and Technology*, vol. 52 (13), 1157-1168.
- Thelwall, M., y Aguillo, I. F. (2003): La salud de las web universitarias españolas. *Revista española de documentación científica*, vol. 26 (3), 291-305.
- Usher, A., y Savino, M. (2006): *A World of Difference: A Global Survey of University League Tables*. Toronto, Ont; Educational Policy Institute, 63.