

El empleo de la Webmetría para el análisis de los indicadores de desempeño y posición financiera de la empresa: un análisis exploratorio en diversos sectores económicos de los Estados Unidos

THE USE OF WEBOMETRICS FOR FINANCIAL PERFORMANCE ANALYSIS AND FINANCIAL POSITION INDICATORS: EXPLORATORY STUDY OF DIFFERENT ECONOMIC SECTORS IN THE UNITED STATES.

Through the development of the Internet over the past two decades, a new discipline that allows the study of the Web from a quantitative point of view has arisen: webometrics. Applying webometrics techniques to management studies, points to the possible existence of a relationship between the number of inlinks that a corporate website receives, and certain economic and financial variables. Bearing in mind that the emerging evidence found is limited to companies from the technological sector, that indeed by its own nature has a strong presence on the Web, this exploratory study provides an extension in the analysis of webometrics to various economic sectors in the United States, confirming the existence of a significant relationship between the exposed variables, and thus opening new possibilities for research in the field of management by using Web-link information.

KEYWORDS: Webometrics, financial performance, data mining, new technologies, economic analysis.

L'UTILISATION DE LA WEBOMÉTRIE DANS L'ANALYSE DES INDICATEURS DE PERFORMANCE ET DE LA POSITION FINANCIÈRE DE L'ENTREPRISE: UNE ANALYSE EXPLORATOIRE DANS DIVERS SECTEURS ÉCONOMIQUES DES ÉTATS-UNIS

RÉSUMÉ: Le développement d'Internet dans les deux dernières décennies a donné lieu à une nouvelle discipline qui permet d'étudier le web d'un point de vue quantitatif: la webométrie. L'application de techniques webométriques à l'étude des entreprises concerne la possible existence de relations entre le nombre de liens que reçoit un site web corporatif et certaines variables économique-financières. À ce jour, les premières évidences dont on dispose se limitent à des entreprises appartenant à des secteurs technologiques qui, par leur nature même, ont une forte présence sur la Toile. La présente étude, de caractère exploratoire, étend l'analyse à d'autres secteurs économiques des États-Unis; elle confirme l'existence d'une relation significative entre les variables mentionnées. Ce qui ouvre de nouvelles possibilités de recherche en matière entrepreneuriale à partir de l'exploitation de l'information tirée de la structure des liens du web.

MOTS CLÉS: Webométrie, performance financière, prospection de données (data mining), nouvelles technologies, analyse économique.

O EMPREGO DA WEBOMETRIA PARA A ANÁLISE DOS INDICADORES DE DESEMPENHO E POSIÇÃO FINANCEIRA DA EMPRESA: UMA ANÁLISE EXPLORATÓRIA EM DIVERSOS SETORES ECONÔMICOS DOS ESTADOS UNIDOS

RESUMO: o desenvolvimento da internet ao longo das duas últimas décadas deu lugar a uma nova disciplina que permite estudar a web a partir de um ponto de vista quantitativo: a webometria. A aplicação de técnicas webométricas ao estudo de empresas aponta para a possível existência de relações entre o número de links que recebe um site corporativo e determinadas variáveis econômico-financieiras. A incipiente evidência encontrada limita-se, em todo caso, a empresas pertencentes a setores tecnológicos que, por sua própria natureza, contam com uma forte presença na Rede. O presente trabalho, de carácter exploratório, estende a análise a diversos setores econômicos dos Estados Unidos, confirmando a existência de uma relação significativa entre as variáveis expostas. Isso abre novas possibilidades de pesquisa no âmbito empresarial a partir do emprego da informação extraída da estrutura de links da web.

PALAVRAS-CHAVE: Webometria, desempenho financeiro, mineração de dados, novas tecnologias, análise econômica.

CORRESPONDENCIA: Esteban Romero Frías. Departamento de Economía Financiera y Contabilidad. Facultad de Ciencias Económicas y Empresariales. Universidad de Granada. Campus Universitario de Cartuja, s/n. C.P. 18071 Granada (España).

CITACIÓN: Romero Frías, E., Vaughan, L., & Rodríguez Ariza, L. (2015). El empleo de la Webmetría para el análisis de los indicadores de desempeño y posición financiera de la empresa: un análisis exploratorio en diversos sectores económicos de los Estados Unidos. *Innovar*, 25(56), 99-112. doi: 10.15446/innovar.v25n56.48993.

ENLACE DOI: <http://dx.doi.org/10.15446/innovar.v25n56.48993>.

CLASIFICACIÓN JEL: C80, D80, E44.

RECIBIDO: Junio 2011, **APROBADO:** Junio 2013.

Esteban Romero Frías

Ph.D. en Contabilidad

Universidad de Granada

Granada, España

Correo electrónico: erf@ugr.es

Liwen Vaughan

Ph.D. en Information Science

University of Western Ontario

London, Ontario, Canadá

Correo electrónico: lvaughan@uwo.ca

Lázaro Rodríguez Ariza

Ph.D. en Ciencias Económicas y Empresariales

Universidad de Granada

Granada, España

Correo electrónico: lazaro@ugr.es

RESUMEN: El desarrollo de Internet a lo largo de las dos últimas décadas ha dado lugar a una nueva disciplina que permite estudiar la Web desde un punto de vista cuantitativo: la webmetría. La aplicación de técnicas webométricas al estudio de empresas apunta a la posible existencia de relaciones entre el número de enlaces que recibe un sitio web corporativo y determinadas variables económico-financieras. La incipiente evidencia encontrada se limita, en todo caso, a empresas pertenecientes a sectores tecnológicos que, por su propia naturaleza, cuentan con una fuerte presencia en la Red. El presente trabajo, de carácter exploratorio, extiende el análisis a diversos sectores económicos de los Estados Unidos, confirmando la existencia de una relación significativa entre las variables expuestas. Esto abre nuevas posibilidades de investigación en el ámbito empresarial a partir del empleo de la información extraída de la estructura de enlaces de la Web.

PALABRAS CLAVE: Webmetría, desempeño financiero, minería de datos, nuevas tecnologías, análisis económico.

Introducción

La importancia de Internet y de la World Wide Web (Web, en adelante), así como la relevancia de la presencia de las empresas en ella, no precisa justificación en nuestros días. Valga como muestra de la dimensión del fenómeno la siguiente noticia ofrecida por Google a mediados de 2008, según la cual el número de páginas indexadas en su base de datos habría superado el

billón de URL¹ únicas (Google blog, 2008). Desde la creación de la Web hace poco más de 20 años, su tamaño no ha dejado de crecer, siendo extremadamente difícil ofrecer una estimación fiable de su dimensión real. La ingente cantidad de información existente recoge manifestaciones de todos los tipos de actividades humanas. La Web contiene reflejos de actividades desarrolladas fuera de ella (*offline*), pero también ha permitido que se desarrollen manifestaciones de carácter genuino, es decir, que sin ella no existirían. La superposición entre lo *offline* y lo *online* se ha visto intensificada de manera significativa en los últimos años, con el surgimiento y popularización de la Web 2.0 (O'Reilly, 2005), la cual ha permitido una democratización en el acceso a los medios de información, eliminando barreras técnicas y económicas para la producción de contenidos por parte de los ciudadanos.

La gran amalgama de información disponible nos permite considerar la Web como una enorme base de datos de naturaleza heterogénea, no estructurada y distribuida. En tanto qué base de datos, podríamos valorar la aplicación de procedimientos destinados a identificar patrones comunes en la información. A lo largo de las dos últimas décadas se han realizado diversos intentos desde distintas áreas de conocimiento. Así, se habla de minería de datos web (*Web Data Mining*), principalmente desde el campo de la informática, o de webmetría, desde el campo de las ciencias de la información. Nuestro trabajo se enmarca en la perspectiva webométrica.

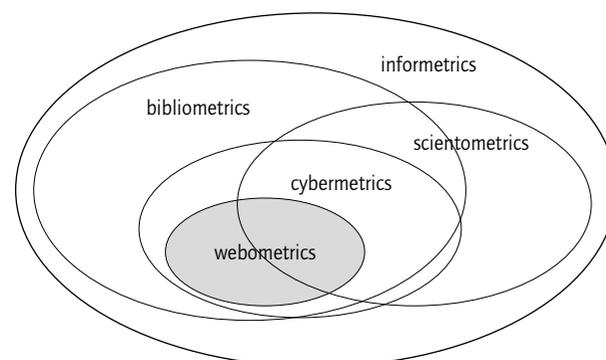
La webmetría y su aplicación al ámbito empresarial

De acuerdo con Björneborn (2004), la webmetría (en inglés se escribe usualmente *Webometrics*) se define como el estudio de los aspectos cuantitativos de la construcción y empleo de recursos, estructuras y tecnologías de la Web sobre la base de una perspectiva bibliométrica. La definición deja clara la relación directa con la bibliometría, permitiendo entender tanto los métodos empleados como los campos de investigación en los que se ha desarrollado principalmente hasta la fecha. Thelwall, Vaughan y Björneborn (2005) simplifican la definición señalando que la webmetría es "*the quantitative study of Web-related phenomena*" (p. 81), incluyendo investigaciones que van más allá de las ciencias de la información. En la misma línea se pronunció

Thelwall (2008). Aunque la webmetría bebe en las fuentes de la bibliometría, debe ser capaz, como manifiestan Thelwall *et al.* (2005), de desarrollar herramientas y técnicas que permitan identificar, observar y cuantificar los cambios sociales que se manifiestan en la Web. En este sentido, la investigación llevada a cabo en los últimos tiempos ha puesto en solfa el carácter profundamente abierto e interdisciplinar de la webmetría (Arroyo, Ortega, Pareja, Prieto y Aguillo, 2005), proporcionando métodos de investigación que pueden ser de provecho para el estudio de fenómenos y problemas propios de otras áreas de conocimiento.

El Gráfico 1 sitúa la webmetría en el contexto de las ciencias de la información (Björneborn e Ingwersen, 2004). La informetría, desde un punto de vista más amplio, analiza todo tipo de información, centrándose principalmente en los flujos de información, sistemas de acceso y recuperación en cualquier forma y soporte. La bibliometría tiene por objeto de estudio documentos en formato impreso (libros, artículos, etc.). De los conceptos y técnica que la bibliometría ha desarrollado se nutren directamente la ciencia-metría, la cibermetría y la webmetría. Prestamos especial atención a las dos últimas por centrarse en el espacio de lo virtual: mientras que la cibermetría es un concepto más amplio que hace referencia al estudio de la información en el ciberespacio, la webmetría se limita principalmente a la Web, que constituye el espacio más relevante de desarrollo de interacciones sociales en red y por tanto la principal fuente de información para la investigación aplicada a las ciencias sociales. Pese a su origen, la investigación llevada a cabo en los últimos años ha subrayado su carácter profundamente interdisciplinar, proporcionando métodos de investigación para la Web, en tanto objeto de estudio, que pueden ser de provecho en diversas áreas de conocimiento, por ejemplo, de manera destacada en los estudios de empresa, entre otros.

GRÁFICO 1. Esquema de la situación de la webmetría en el contexto de las ciencias de la información



Fuente: Adaptado a partir de Björneborn e Ingwersen (2004, p. 1217).

¹ URL son las siglas en inglés de *Uniform Resource Locator*, que se puede traducir por "localizador uniforme de recurso". Se trata de una secuencia de caracteres, de acuerdo a un formato estándar, que se emplea para identificar recursos, como documentos e imágenes en Internet, a través de su localización.



- La investigación basada en palabras clave: realizando análisis de impacto de ideas, conceptos, marcas, organizaciones, empresas, etc., a partir del recuento y análisis de las páginas web proporcionadas por los motores de búsqueda al formular una consulta mediante palabras clave.
- La investigación basada en los hiperenlaces: destacando, por una parte, el análisis de impacto de enlaces y, por otra, la representación gráfica de relaciones a través de enlaces. El primero permite valorar la intensidad de un fenómeno o la relevancia de una entidad en función del número de enlaces recibidos; el segundo pone esta herramienta al servicio del estudio de las relaciones entre diversas páginas web, proporcionando un instrumento de análisis de redes sociales.

Una tercera vía se basa en el comportamiento de los usuarios al realizar búsquedas en páginas web (Lambert, 2008; Ravid, Bar, Rafaeli y Baruchson, 2007) o en buscadores comerciales, particularmente a través del uso de Google Trends. Hay una literatura cada vez más rica en este sentido aplicada a empresas, como veremos posteriormente.

Frente a otras posibles variables webmétricas (Faba, Guerrero y Moya, 2004), la empleada en este artículo: la de los hiperenlaces, presenta como principal ventaja la precisión, ya que evitan por ejemplo el ruido que genera el empleo de términos con múltiples significados en el caso de la investigación basada en palabras clave. Un hiperenlace (o enlace, en adelante) constituye una referencia o un elemento de navegación, contenido en un documento que conduce hacia otra sección del mismo documento o hacia otros documentos completamente distintos. Los enlaces representan la estructura invisible de la Web, conectando páginas que de otra manera permanecerían aisladas y difícilmente accesibles (Berners-Lee, 1999). Se podrían considerar como un voto a favor de la página a la que enlazan, indicando la relevancia de la información contenida en ella. Quizá el ejemplo paradigmático del empleo de los hiperenlaces como indicador de relevancia es el *Pagerank* de Google (Brin y Page, 1998), sistema empleado por el buscador para la clasificación de los resultados contenidos en su base de datos una vez que el usuario realiza una búsqueda. El *Pagerank*, aparte de otros factores, asume que las páginas que reciben un mayor número de enlaces (y

también aquellas que reciben enlaces de páginas más importantes) son las más relevantes y útiles para el usuario. La efectividad de este sistema ha hecho que actualmente la mayoría de buscadores comerciales basen sus algoritmos de clasificación de resultados en la estructura de enlaces de la Web (Batelle, 2005; Machill, Beiler y Zenker, 2008). En una economía basada en la información, el posicionamiento web (Batelle, 2005) se convierte en un factor clave para el éxito empresarial condicionando la toma de decisiones. La presente investigación ahonda en esta línea haciendo un llamado de atención a las organizaciones para que analicen la estructura de enlaces que vincula su web con otras.

De la bibliometría a la webmetría: el análisis de hiperenlaces

La evidente analogía entre páginas web e hiperenlaces, por un lado, y artículos científicos y citas bibliográficas, por otro, hace que desde la creación de la Web hayan surgido intentos de aplicar los conceptos y herramientas de la bibliometría a ella. La idea de que un enlace a una página Web puede interpretarse como un voto a favor de la misma se deriva del funcionamiento de los sistemas de citación y de elaboración de clasificaciones de impacto en el campo de la investigación científica (Garfield, 1979).

En el contexto de la Web se han puesto de manifiesto relaciones entre variables como son los enlaces recibidos o *inlinks* (número de enlaces que apuntan a una determinada URL o dominio) y diversas medidas de desempeño de la actividad investigadora. Por ejemplo, se han estudiado las relaciones entre enlaces recibidos por sitios web de universidades con los niveles de investigación alcanzados por las mismas (Thelwall, 2001; Smith y Thelwall, 2002; Tang y Thelwall, 2003; Aguillo, Granadino, Ortega y Prieto, 2006; Aguillo, Ortega, Prieto y Granadino, 2007), enlaces a sitios web de facultades con su productividad investigadora (Chu, He y Thelwall, 2002) y enlaces a sitios web de revistas con la calidad de las revistas (Vaughan y Hysen, 2002; Vaughan y Thelwall, 2003). Otros trabajos han empleado los enlaces para medir la visibilidad y el impacto de un grupo de sitios web (Ingwersen, 1998; Vreeland, 2000; Chu, He y Thelwall, 2002), para analizar patrones de comunicación informal entre académicos (Li, 2003; Wilkinson, Harries, Thelwall y Price, 2003; Tang y Thelwall, 2004), para analizar la relación con patrones geográficos o distancias entre organizaciones (Thelwall, 2002), o para identificar áreas académicas con un mayor impacto en la Web (Thelwall, Vaughan, Cothey, Li y Smith, 2003). Muchos de los trabajos realizados en el campo del estudio de la ciencia son de carácter exploratorio, buscando poner

de manifiesto relaciones entre fenómenos observados en la Web y fenómenos propios del mundo físico. Esta exploración progresiva es la que ha permitido sentar las bases para el desarrollo de sistemas cada vez más elaborados que explotan la información contenida en los hiperenlaces.

Como apunta Holmberg (2009), si se tiene en cuenta la naturaleza en apariencia caótica de la Web, es muy significativo el hecho de que, en determinados ámbitos, los estudios webmétricos hayan encontrado correlaciones significativas entre enlaces recibidos y otras variables de reconocido valor, ajenas a la Web; por ejemplo, la calidad de una publicación científica o determinadas variables financieras.

Estudios webmétricos de empresas

El estudio webmétrico de sitios web comerciales no ha alcanzado el grado de desarrollo de sitios académicos. Esto resulta paradójico si tenemos en cuenta el predominio de los primeros y el incremento continuo de la importancia del comercio electrónico (Thelwall *et al.*, 2005; Vaughan y Wu, 2004). La explicación reside fundamentalmente en la inercia investigadora de esta nueva disciplina, en tanto que procede de la bibliometría.

Diversos trabajos han formulado un análisis cuantitativo de la Web en el campo de empresas, utilizando para ello métodos que podrían encuadrarse dentro de la webmetría o de la minería de datos (Romero, 2010). Por ejemplo, Liu, Ma y Yu (2001) desarrollaron un sistema para obtener información a partir de sitios web de empresas competidoras. Algunos trabajos (por ejemplo, Thelwall, 2000; Wormell, 2001; Reid, 2003) abordaron el estudio de sitios web de empresas, centrándose principalmente en sus contenidos y su diseño. Posteriormente, algunos trabajos (Chau, Shiu, Chan y Chen, 2007; Vaughan y You, 2009) han empleado, para la investigación en empresas, combinaciones de hiperenlaces y de palabras clave. Tras la popularización de la Web 2.0, algunos estudios (por ejemplo, Kirkpatrick, Roth y Ryan, 2005; Das y Sisk, 2005) han examinado la importancia de los blogs para la obtención de información empresarial. Chen, Hu y Liu (2007) analizaron las relaciones entre la actividad desplegada en los blogs corporativos de un conjunto de empresas del *ranking* Fortune 500 y su desempeño financiero.

El estudio de las búsquedas realizadas por los usuarios en buscadores comerciales, en concreto a través del servicio Google Trends (o Google Search Insights), ha perseguido la realización de predicciones económicas. Choi y Varian (2009, 2012) muestran cómo el empleo de estos indicadores permite predecir varios indicadores relacionados con el desempleo, la demanda de automóviles y el turismo.

Radinsky, Davidovich y Markovitch (2009), Huang y Penna (2009), Kholodilin, Podstawski y Siliverstovs (2010), y Vosen y Schmidt (2011) han examinado el uso de los datos de búsquedas para medir el sentimiento del consumidor y las ventas. En el sector inmobiliario, Wu y Brynjolfsson (2010) encontraron evidencia de que los datos de búsquedas pueden proporcionar información sobre los precios y el volumen de transacciones inmobiliarias futuros.

En el campo financiero, Preis, Reith y Eugene (2010) encontraron correlaciones entre los volúmenes de transacciones de las empresas del S&P 500 y el volumen de búsquedas de los nombres de las empresas. Por su parte, Da, Engelberg y Gao (2011) utilizaron los datos de Google como un indicador de la atención que los inversores prestan a acciones y las fluctuaciones en el precio.

Si bien estos estudios han dado resultados positivos, las conclusiones que se pueden extraer de ellos no son tan prometedoras en cuanto a su capacidad predictiva como se habría esperado. En línea con lo también manifestado por Ginsberg, Mohebbi, Patel, Brammer, Smolinski y Brilliant (2009), más que de predicciones podríamos hablar de una anticipación de datos que generalmente se demora a través de los cauces oficiales: se gana pues en inmediatez.

Centrándonos en la investigación basada en hiperenlaces, Vaughan y Romero (2010) han realizado un análisis del sector bancario internacional mediante la combinación de distintos métodos webmétricos; mientras que Vaughan, Tang y Du (2010) se han centrado en la elaboración de perfiles de empresas a partir de la información obtenida a través de diversas fuentes en la Web.

En este conjunto de trabajos existe una línea que vincula el número de hiperenlaces que reciben los sitios web de empresas con variables de posición y de desempeño financieros, en la línea de lo que se ha realizado en el campo bibliométrico con medidas de la producción científica. Ello hace que sea necesaria la elaboración de trabajos de carácter exploratorio que permitan determinar la validez de estos planteamientos en su aplicación a empresas. Solo mediante una evaluación preliminar de la realidad económica a través de los procedimientos webmétricos será posible emplear esta nueva fuente de información para la elaboración de modelos que permitan predecir o anticipar variables financieras, confirmar la evolución económica de determinados sectores o fortalecer sistemas de análisis empresarial, entre otros posibles empleos que la comunidad científica o empresarial pueda encontrar. Es especialmente relevante la potencial utilización de esta fuente de información en el análisis de activos intangibles, así como de la evolución empresarial hacia modelos de negocio más basados en la información (Romero y Vaughan, 2010).

Dentro de la variedad de estudios expuestos, una serie de trabajos han explorado las relaciones entre el número de enlaces recibidos por sitios web de empresas y sus variables de posición y de desempeño financieros. El objetivo de estos trabajos, de carácter exploratorio, ha sido el de sentar las bases para un posible empleo de este tipo de variable web como indicador de desempeño empresarial o de la dimensión o envergadura de la entidad. Hasta ahora, las investigaciones se han circunscrito, en cuanto a la actividad, al sector de tecnologías de la información y la comunicación, y, en cuanto al ámbito geográfico, a los mercados norteamericano (Estados Unidos y Canadá) y chino. Así, Vaughan y Wu (2004) exploraron la existencia de correlaciones entre algunas variables económico-contables y los enlaces que reciben empresas del mercado chino. Encontraron correlaciones significativas entre el número de enlaces recibidos por las 100 mayores empresas de tecnologías de la información y variables como: ingresos brutos (0,51**), resultado (0,30**) y gasto en I+D (0,64**), con coeficientes de correlación significativos al nivel 0,01.

Vaughan (2004a) extendió el anterior trabajo a empresas del sector de tecnologías de la información en China y en Estados Unidos, con resultados positivos. Vaughan (2004b) compara las empresas de tecnologías de la información de Canadá con las de Estados Unidos. Por otra parte, Vaughan y Romero (2010) observan correlaciones significativas en un estudio del sector bancario internacional.

Los anteriores resultados invitan a pensar que los enlaces recibidos por sitios web de empresas podrían emplearse como un indicador complementario de la dimensión o del desempeño de las mismas. Dada la correlación positiva existente entre empresas pertenecientes a un mismo sector, se podría afirmar que las entidades con un mejor desempeño financiero tienen también una presencia en la Web más intensa, atrayendo mayor atención de otras páginas que consecuentemente establecen vínculos con ellas.

Justificación y objetivos del trabajo

El presente estudio amplía la evidencia de los anteriores trabajos respondiendo a la siguiente pregunta: ¿existen relaciones significativas entre variables financieras y la variable enlaces recibidos en sectores de actividad de naturaleza diversa, al margen del sector de las tecnologías de la información? Hasta ahora, el sector de actividad se ha considerado una variable determinante a la hora de explicar los resultados obtenidos, lo cual ha hecho que la investigación previa se haya centrado fundamentalmente en sectores tecnológicos, más propensos a tener una mayor presencia en la Web. Si en sectores de actividad con un componente tecnológico muy diferente y con distintos modelos de negocio

en relación con la Web se mantienen estos vínculos, podríamos plantear el desarrollo de modelos explicativos de carácter general que incluyeran este tipo de variable web. Así pues, el objetivo de la investigación se materializa en una serie de hipótesis que se evaluarán para cada uno de los sectores de actividad contemplados:

H1: Existe una relación directa entre el número de enlaces recibidos y variables de posición financiera.

H2: Existe una relación directa entre el número de enlaces recibidos y variables de desempeño financiero en términos absolutos.

H3: Existe una relación directa entre el número de enlaces recibidos y variables de desempeño financiero en términos relativos.

H4: Existe una relación directa entre el número de enlaces recibidos y variables financieras correspondientes a un conjunto de empresas de actividades heterogéneas.

Método

Empresas y sectores estudiados

La elección de Estados Unidos como país de referencia para nuestra investigación se debe a las siguientes razones:

- Es un país donde la presencia comercial de las empresas en la Web es muy elevada, además de ser un país pionero en el desarrollo de nuevas tendencias en el campo tecnológico, especialmente de Internet.
- Se trata de la principal economía mundial, con lo que es más probable que las empresas cuenten con la dimensión y envergadura suficientes para tener en la Web un fiel reflejo de sus actividades.
- Se trata de uno de los países que ha sido estudiado para el caso de empresas de tecnologías de la información en trabajos previos (Vaughan, 2004a, 2004b), por lo que permite establecer comparaciones con los resultados obtenidos anteriormente.

Las empresas estudiadas cotizan en bolsa, dado que son las obligadas a publicar su información financiera en Estados Unidos. Los sectores seleccionados son los siguientes: Bancos (*Commercial banks*), Construcción (*Construction of buildings*), Grandes almacenes (*General merchandise stores*), Minería (*Mining*) y Servicios públicos (*Utilities*). La selección pretende considerar sectores con actividades y modelos de negocio diversos, de modo que se pueda comprobar si en todos ellos existen correlaciones significativas y qué nivel de correlación existe. Un reciente trabajo (Romero y Vaughan, 2010), basado en el análisis de

co-enlaces, apunta la existencia de diversos patrones de presencia en la Web en función del sector de actividad. Al mismo tiempo, se ha intentado que los sectores escogidos contaran con suficiente número de empresas para los análisis estadísticos.

Adicionalmente, hemos incluido las empresas que forman el índice *Dow Jones Industrial*, con el fin de comprobar la hipótesis 4. Vaughan y Wu (2004) apuntan a que únicamente en conjuntos de empresas que realizan la misma actividad y, por tanto, están sujetos a modelos de negocio similares y a una exposición a Internet parecida se detectan correlaciones significativas entre número de enlaces y variables económico-financieras.

La Tabla 1 describe la muestra de empresas, obtenida a partir de la base de datos Mergent². La primera columna incluye el total de empresas activas en la base de datos para el código NAICS indicado (sistema de clasificación de actividades comerciales en Estados Unidos). Siempre que el número de empresas activas fuera superior a 100, se ha tomado una muestra aleatoria de este tamaño, suficiente para realizar las pruebas estadísticas. En el caso de que fuera inferior se ha tomado el conjunto de empresas activas disponibles. Posteriormente, se ha procedido a excluir a aquellas que no cumplen con las siguientes condiciones:

- Contar con una página web única, es decir, no compartir una dirección web con empresas pertenecientes a un mismo grupo, por ejemplo.
- Disponer de datos de activos totales para el ejercicio 2007.

Por último, una vez efectuadas las correspondientes exclusiones, se indica el número final máximo de empresas válidas de la muestra para cada sector.

La selección de empresas se realiza en función de los datos disponibles en la base de datos en enero de 2009. Dicha selección condiciona la obtención de los datos de enlaces recibidos. Posteriormente, en junio de 2010, se han ampliado los datos financieros de las empresas seleccionadas incluyendo los ejercicios 2008 y 2009. Es preciso indicar que la composición del índice *Dow Jones Industrial* corresponde al 4 de diciembre de 2008³. En junio de 2009, Citigroup y General Motors salen del índice y entran Cisco Systems y Travelers Companies. Por ello, se ha decidido eliminar las dos primeras compañías, que se incluían en la selección inicial, quedando únicamente 28 compañías en el grupo, frente a las 30 que lo componen.

² <http://www.mergent.com>

³ <http://www.djindexes.com/mdsidx/?event=showAverages>

TABLA 1. Empresas incluidas en el estudio

Sectores comerciales (código NAICS)	Total de empresas "activas"	Tamaño de la muestra aleatoria	Exclusión por falta de página web	Exclusión por falta de información financiera	Número final de empresas válidas
Bancos (52.211)	581	100	6	12	82
Construcción (236)	57	57	16	19	22
Grandes almacenes (452)	35	35	6	6	23
Servicios públicos (221)	275	100	18	14	68
Minería (21)	671	100	21	27	52
Dow Jones Industrial	30	30	-	2	28

Fuente: Elaboración propia.

Obtención de los datos de empresas

En sintonía con las hipótesis formuladas, las variables financieras seleccionadas son representativas de los tres siguientes tipos de variables:

- Variables de posición financiera: activos totales.
- Variables de desempeño financiero (en términos absolutos): resultado neto, ingresos, EBITDA.
- Variables de desempeño financiero (en términos relativos): ROA (*Return on Assets*), ROE (*Return on Equity*).

Al margen, se ha incluido también la variable "número de empleados" como indicador del tamaño de la empresa. Los datos se han obtenido para los ejercicios 2007, 2008 y 2009. Si tenemos en cuenta que los datos de enlaces recibidos se corresponden a enero de 2009, asumimos que recogen toda la información generada durante el ejercicio 2007 y la mayoría de la información generada en 2008. La información financiera de 2009 nos permitiría indicar si la información de enlaces recibidos puede funcionar como un indicador anticipado de las variables financieras.

Obtención de los datos de enlaces recibidos

Una vez identificadas las empresas que se van a analizar, se utiliza Google para determinar sus sitios web. En aquellos casos en los que una empresa cuenta con varios dominios se selecciona aquel que mayor número de enlaces recibe. Posteriormente, para obtener los datos de enlaces recibidos, se emplea Yahoo!, ya que en la fecha de recuperación de la información es el buscador comercial que facilita las mejores opciones para el diseño de búsquedas complejas y el que proporciona los resultados más completos para este tipo de búsquedas, por encima de Google y MSN (Bing, en la actualidad).

El empleo de buscadores comerciales para la obtención de datos presenta algunas limitaciones importantes,

especialmente si se tiene en cuenta que son herramientas no diseñadas para la investigación académica sino con fines comerciales (Bar-Ilan, 2001; Faba *et al.*, 2004). Entre los problemas que presentan podemos señalar los siguientes:

- La volatilidad en los resultados, que pueden cambiar en cortos periodos de tiempo o en función de la localización geográfica.
- La invisibilidad de buena parte de la Web debido a cuestiones técnicas en el diseño de las páginas y de las bases de datos, que impide que los buscadores puedan indexar toda la Web.
- Los resultados basados en estimaciones, debido a que sus fines comerciales hacen que prime la rapidez y la eficiencia al ofrecer resultados frente a la exhaustividad; entre otras cuestiones.

Sin embargo, el perfeccionamiento de los buscadores ha hecho que los resultados que ofrecen sean cada vez más estables (Vaughan, 2004c). Al mismo tiempo, cabe destacar que estas herramientas son la única fuente de información viable de la cual disponemos para obtener determinadas variables como, por ejemplo, los enlaces recibidos. La validez de sus datos queda respaldada por una amplia labor investigadora efectuada desde el campo de las ciencias de la información.

En nuestro trabajo se han empleado dos consultas distintas:

- linkdomain:abc.com –site:abc.com
- link:http://www.abc.com –site:abc.com

La primera está basada en el operador "linkdomain", que proporciona todos aquellos enlaces (URL) que apuntan a cualquier página perteneciente al dominio especificado. La segunda, por su parte, emplea el operador "link", que proporciona las páginas que enlazan únicamente a la dirección URL indicada. Los resultados ofrecidos en el primer caso son más numerosos, ya que incluyen los enlaces que

TABLA 2. Datos descriptivos de la variable enlaces recibidos

Variable "enlaces recibidos"	N	Mínimo	Máximo	Mediana	Media	Desviación típica
Bancos	82	34	22.500	553	1.772,57	3.614,27
Construcción	22	62	158.000	922,5	10.274,05	33.335,24
Grandes almacenes	23	422	2.800.000	14.200	234.355,43	624.770,13
Servicios públicos	68	7	26.200	1.445	4.178,60	5.251,28
Minería	52	12	14.900	291	919,46	2.476,97
Dow Jones Industrial	28	42.800	45.500.000	257.500	2.483.678,57	8.534.851,18

Fuente: Operador "linkdomain", febrero de 2009.

apuntan a la URL principal de la página de la empresa más todos aquellos que apuntan a URL dentro de directorios del dominio o de subdominios (por ejemplo, http://def.abc.com). Si bien los resultados que producen ambos operadores son similares, en el trabajo se utiliza como referencia el operador "linkdomain", ya que conceptualmente recoge las referencias hechas por otras páginas a la web de una empresa de forma más completa. Al margen, el componente "-site:abc.com" de la fórmula de consulta tiene por objeto eliminar aquellos resultados que proceden del mismo dominio, es decir, auto-referencias, las cuales, al ser creadas por la propia empresa, pueden conllevar un sesgo. Los datos se recogen, empleando el operador "linkdomain", en dos momentos distintos: a principios de febrero de 2009 y en mayo de 2009. Adicionalmente, en mayo de 2009 también se recogen datos empleando el operador "link".

Resultados

Dado que la mayoría de las variables en el estudio, especialmente las correspondientes a enlaces recibidos, presentan una distribución muy asimétrica, no satisfaciendo el supuesto de normalidad, se opta por utilizar el coeficiente de correlación de Spearman⁴ con el fin de determinar si existen relaciones significativas entre la variable de enlaces recibidos y las variables económico-financieras. En la Tabla 2 se incluyen datos descriptivos de la variable enlaces recibidos para cada uno de los grupos de empresas representados en el trabajo. En primer lugar, debemos destacar que se trata de datos con una distribución

asimétrica, en la que por lo general un pequeño número de empresas concentran la mayor parte de los enlaces recibidos. El sector con mayor número de enlaces es el de "grandes almacenes", que incluye empresas de distribución al por menor, las cuales a lo largo de los últimos años han visto incrementada su presencia en la Web a partir de la incorporación de catálogos online y de potentes líneas de comercio electrónico. En el lado contrario, el sector con menor número de enlaces es el de la minería, un sector con una baja exposición al cliente final. Las empresas que componen el índice Dow Jones Industrial reciben un volumen de enlaces por encima del resto. Esto es lógico en tanto que hablamos de una selección de las empresas más importantes de Estados Unidos.

A continuación, prestamos atención a las distintas mediciones efectuadas de la variable enlaces recibidos. De los tres conjuntos de datos, los resultados obtenidos mediante el operador "linkdomain" proporcionan una correlación significativa y más alta que los derivados del empleo de "link". Por otra parte, salvo excepciones, en la mayoría de los casos las correlaciones con los enlaces de febrero son más elevadas que con los datos de mayo. Los coeficientes de correlación entre los distintos conjuntos de datos de enlaces recibidos son muy elevados, lo cual indica que los resultados que proporcionan los buscadores son consistentes y relativamente estables en el tiempo.

La Tabla 3 muestra los coeficientes de correlación entre los enlaces recibidos (recogidos en febrero de 2009) y las variables económico-financieras consideradas en el trabajo. Se han adoptado como referentes para el trabajo los enlaces obtenidos en febrero por tres razones: 1) el empleo del operador "linkdomain" es conceptualmente más apropiado para los objetivos de la investigación, ya que incluye todos los enlaces a cualquier página dentro del sitio web; 2) la medición está próxima al cierre del ejercicio 2008 con

⁴ El coeficiente de correlación de Spearman, ρ (rho), es una medida de la correlación no paramétrica entre dos variables que se emplea como alternativa al coeficiente de correlación de Pearson cuando no se cumplen las condiciones exigidas para su utilización (Vaughan, 2009).

TABLA 3. Coeficientes de correlación de Spearman para cada uno de los grupos de empresas considerados

	Año	Bancos	Construcción	Minería	Grandes almacenes	Servicios públicos	Dow Jones Industrial
Empleados	2009	0,714**	0,562**	0,509**	0,885**	0,826**	0,346
	2008	0,718**	0,620**	0,482**	0,868**	0,826**	0,363
	2007	0,729**	0,408	0,490**	0,842**	0,839**	0,327
Activos	2009	0,746**	0,825**	0,683**	0,889**	0,814**	0,253
	2008	0,749**	0,811**	0,578**	0,869**	0,820**	0,285
	2007	0,721**	0,816**	0,485**	0,868**	0,822**	0,315
Resultado Neto	2009	0,021	-0,616**	-0,133	0,479*	0,741**	0,380*
	2008	0,683**	0,791**	0,504**	0,865**	0,819**	0,316
	2007	0,667**	0,607**	0,567**	0,873**	0,792**	0,284
Ingresos	2009	0,754**	0,849**	0,468**	0,912**	0,812**	0,270
	2008	0,753**	0,860**	0,489**	0,897**	0,809**	0,260
	2007	0,729**	0,886**	0,525**	0,886**	0,791**	0,230
EBITDA	2009	-	-0,681**	0,388**	0,546**	0,755**	0,408*
	2008	-	-0,636**	-0,005	0,857**	0,710**	0,428*
	2007	-	-0,252	0,289*	0,944**	0,786**	0,372
ROA	2009	0,039	-0,158	0,556**	0,063	0,252	0,220
	2008	0,245*	0,286	-0,079	0,319	0,000	0,144
	2007	0,175	-0,182	-0,142	0,307	-0,055	0,045
ROE	2009	0,035	-0,179	0,301	0,058	0,204	0,102
	2008	0,228*	0,247	0,029	0,325	-0,106	0,073
	2007	0,198	-0,228	-0,058	0,388	-0,037	-0,042

* Coeficientes de correlación significativos al nivel 0,05.

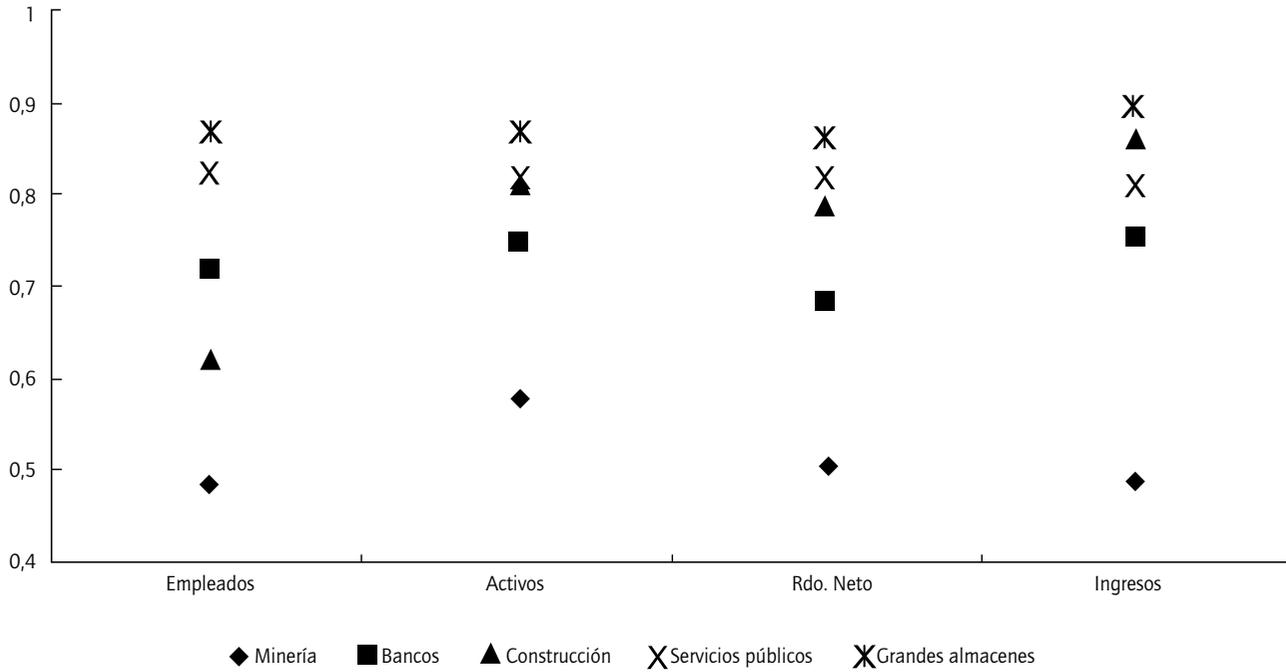
** Coeficientes de correlación significativos al nivel 0,01.

Fuente: Elaboración propia.

lo que permite valorar de manera más estrecha la relación con ese conjunto de datos, y 3) los coeficientes de correlación que arroja son más elevados que los correspondientes a la medición realizada en mayo. En las filas de la Tabla se incluyen, para cada variable, los datos correspondientes a los ejercicios 2007, 2008 y 2009. Pretendemos explorar en cuál de los tres ejercicios los coeficientes de correlación son más elevados. En principio podemos considerar que toda la información generada en el ejercicio 2007 se encontraría ya recogida en la variable "enlaces recibidos", así como toda o la gran mayoría de la información del ejercicio 2008. La inclusión de 2009 pretende valorar su potencial como indicador anticipado de las variables financieras. Los resultados muestran que, en términos generales, los coeficientes de correlación para los distintos años no varían demasiado debido a que las variables contables de un año a otro también están muy correlacionadas. Sin embargo, la situación de crisis que se manifiesta de manera más acusada en las cuentas anuales de 2009 hace que se produzcan variaciones relevantes. En función de lo apuntado anteriormente, vamos a tomar, como referencia principal para el análisis, las variables financieras de 2008.

En relación con la hipótesis 1, se observa la existencia de correlaciones significativas con la variable de posición financiera (activos totales) en todos los sectores empresariales considerados. En relación con la hipótesis 2, las variables de desempeño financiero en términos absolutos (resultado neto, ingresos, EBITDA) presentan también correlaciones significativas, si bien en el año 2009 tanto en el sector de la banca como de la minería no son significativas disminuyendo también el coeficiente de correlación en las otras industrias con respecto al ejercicio anterior. No es el caso de las variables ingresos y EBITDA que se mantienen en niveles parecidos. En relación con la hipótesis 3, las variables de desempeño financiero en términos relativos (ROA y ROE) no presentan correlación significativa alguna. La hipótesis 4 se refiere a la existencia de correlaciones significativas en grupos de empresas heterogéneas (no pertenecientes a un mismo sector empresarial), en este caso las integrantes del índice Dow Jones Industrial. Los datos confirman la no existencia de correlaciones significativas, con la excepción de la variable EBITDA, en un nivel bajo.

GRÁFICO 2. Diagrama que muestra el lugar que ocupa cada sector en cuanto al coeficiente de correlación entre el número de enlaces y las variables económico-financieras indicadas para 2008



Fuente: Elaboración propia.

No se puede afirmar que los coeficientes de correlación con los datos del ejercicio 2008 sean mayores en términos generales que con los de 2007 y 2009.

El Gráfico 2 muestra el coeficiente de correlación entre el número de enlaces recibidos y los datos económico-financieros del ejercicio 2008. Se observa cómo en cada sector el coeficiente de correlación para las distintas variables se sitúa en un nivel similar. Así, se comprueba cómo el sector con correlaciones más altas, en todas las variables, es el de grandes almacenes, mientras que el sector minero presenta las correlaciones más bajas. Los datos sugieren que cuanto mayor es la presencia y la exposición a la Web de las empresas en virtud del tipo de actividad económica que realizan, más relevante es la información que la Web proporciona a través de la variable hiperenlaces recibidos. Ello indica que los métodos webmétricos pueden ser más efectivos para estudiar sectores con modelos de negocio más intensivos en información y en la gestión del conocimiento en la red (Romero y Vaughan, 2010).

Conclusiones, limitaciones y futuras líneas de investigación

Los resultados apuntados indican que en los distintos sectores, por muy diversos que estos sean, existe una vinculación entre el número de enlaces que reciben las páginas

web de las empresas y su dimensión económica medida a través de sus variables de posición y desempeño financieros. El Gráfico 2 muestra cómo cada sector presenta niveles diversos de correlación. Estos niveles podrían explicarse por la naturaleza de la actividad y el modo en que esto condiciona la presencia de las empresas en la Web. Las empresas de grandes almacenes están muy enfocadas al usuario final y, por tanto, es lógico pensar que reciben una mayor atención en este medio que otro tipo de empresas. La orientación al usuario final se traduce en que este sector es el que mayor número de enlaces recibe entre todos los sectores. Cabe también subrayar que la existencia de correlaciones entre variables únicamente se observa entre grupos homogéneos de empresas, como corroboran los datos correspondientes al índice Dow Jones y las investigaciones anteriores (Vaughan y Wu, 2004).

El número de empleados presenta un comportamiento y unos coeficientes de correlación similares a los del volumen de activos en todos los sectores, demostrando ser una variable que pone de manifiesto de forma adecuada la dimensión de la empresa con independencia del sector en que nos encontremos.

La comparación con los resultados de los trabajos anteriormente realizados en el sector de tecnologías de la información (Vaughan, 2004a, 2004b; Vaughan y Wu, 2004) permite destacar que los coeficientes de correlación,

cuando son significativos, se han incrementado. Un factor explicativo sería el gran desarrollo de Internet en los últimos cinco años, especialmente con la aparición de la Web 2.0 (O'Reilly, 2005).

La existencia de correlación entre las dos variables no implica necesariamente la existencia de causalidad entre ellas. Desde el punto de vista de la minería de datos, el hecho de identificar que la variable "número de enlaces recibidos" está correlacionada de forma significativa con variables económico-financieras y es suficiente de cara a obtener información empresarial que pueda ser útil para la toma de decisiones. En todo caso, si bien no es objeto de este artículo el analizar las posibles relaciones causales entre ambas, son variadas las explicaciones que justifican el que empresas con mayor dimensión o mejores resultados tengan también una mayor presencia en Internet (ya sea a nivel de divulgación de información, de actividades de e-commerce, etc.) y, por tanto, reciban también mayor número de enlaces. Cabe precisar que el número de enlaces que se recaba de los buscadores comerciales proviene de la acumulación de enlaces que a lo largo del tiempo se han ido creando en la Web a dicha página o dominio. Tiene pues un carácter acumulativo, que de alguna manera se asemeja a la naturaleza de las magnitudes de posición financiera de la empresa, esto es, aquellas magnitudes que en términos generales sintetizan la dimensión de la entidad desde el momento de su creación hasta el momento presente. Las magnitudes que revelan el desempeño empresarial se calculan de forma periódica, generalmente cada año, por lo que no recogen una visión acumulativa de la empresa que indique la evolución desde su origen. Aún así, los resultados o el volumen de ingresos de una empresa, aunque se calculen para periodos anuales, por lo general siguen una tendencia sostenida año tras año, siendo de esperar, por ejemplo, que el volumen de ingresos de una empresa sea mayor al quinto año de vida que al primero. Es pues coherente pensar que las correlaciones entre número de enlaces y variables económico-financieras serán significativas y más elevadas cuando se comparen magnitudes en términos absolutos (activos, pasivos, volumen de ingresos, resultados) frente a magnitudes en términos relativos (tales como ROA o ROE).

Entre las limitaciones del trabajo podemos identificar algunas provenientes de la elección de empresas de Estados Unidos, lo cual ha obligado a limitar el estudio a entidades cotizadas, que son aquellas para las que existe información económico-financiera disponible. Esto ha hecho que las empresas incluidas sean de gran tamaño y que en determinados sectores el número de entidades considerado en la muestra sea pequeño. En cualquier caso, como comentamos anteriormente, esta elección se ha fundamentado

en que, dado el estado de los estudios previos en esta línea (Vaughan, 2004a, 2004b), hasta el momento Estados Unidos es el país de referencia para establecer comparaciones con los resultados anteriores.

Por otra parte, toda investigación basada en la obtención de información a partir de los buscadores comerciales está limitada por las posibilidades que estos ofrecen, ajenas en muchos casos de la investigación académica. Su empleo, sin embargo, sirve de base a gran parte de la investigación webométrica publicada en revistas de prestigio y supone además una aproximación razonable al vasto conjunto de información que conforma la Web. Un problema adicional asociado con esta fuente de información es la posibilidad de que los buscadores comerciales dejen de ofrecer la posibilidad de realizar este tipo de consultas. Recientemente, Yahoo! y Microsoft han llegado a un acuerdo de fusión de sus servicios de búsqueda, lo cual puede repercutir negativamente en las expectativas de investigación en este sentido.

De cara a la investigación futura, es preciso ampliar la evidencia a otras áreas geográficas y a distintos conjuntos de empresas, por ejemplo, entidades de menor tamaño no cotizadas. Otra línea de trabajo se centraría en estudiar la evolución de los enlaces recibidos por los sitios web de empresas al objeto de poder valorar su potencial capacidad predictiva de indicadores financieros. Por último, conviene valorar el empleo de la variable "enlaces recibidos" como forma de medir los activos intangibles, vinculados a Internet, de las empresas. En el futuro, veremos también cómo otras variables extraídas de servicios Web 2.0 son empleadas para la obtención de información empresarial.

Para concluir, el presente trabajo aporta nueva evidencia sobre la vinculación entre la variable "número de enlaces" recibidos por sitios web de empresas y sus variables financieras. Se ha cubierto la necesidad de ampliar el alcance del análisis a sectores que, por su diferente naturaleza, podrían presentar comportamientos distintos en Internet. Además, se ha trabajado con un mayor número y variedad de variables económico-financieras en comparación con las incluidas en los trabajos anteriores. El trabajo representa además uno de los primeros que, en el ámbito de los estudios de empresa en general, aborda el estudio de la variable "enlaces recibidos" por sitios web de empresas y un conjunto de variables económico-financieras.

Referencias bibliográficas

Aguillo, I.F., Granadino, B., Ortega, J.L., & Prieto, J.A. (2006). Scientific research activity and communication measured with Cybermetrics

- indicators. *Journal of the American Society for Information Science and Technology*, 57(10), 1296-1302.
- Aguillo, I.F., Ortega, J. L., Prieto, J. A., & Granadino, B. (2007). Indicadores Web de actividad científica formal e informal en Latinoamérica. *Revista Española de Documentación Científica*, 30(1): 49-60.
- Arroyo, N., Ortega, J.L., Pareja, V. Prieto, J.A., & Aguillo, I. (2005). "Cibermetría. Estado de la cuestión". En: *Novenas Jornadas Españolas de Documentación*. Madrid: FESABID 2005. 14-15 de abril de 2005.
- Bar-Ilan, J. (2001). Data collection methods on the Web for informetric purposes—A review and analysis. *Scientometrics*, 50(1), 7-32.
- Battelle, J. (2005). *The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture*. London: Portfolio.
- Berners, T. (1999). *Weaving the Web*. San Francisco: Harper.
- Björneborn, L. (2004). *Small-world link structures across an academic Web space: A library and information science approach*. Doctoral dissertation. Royal School of Library and Information Science, Copenhagen, Denmark. Disponible en: <http://vip.db.dk/lb/phd/phd-thesis.pdf> [consultado el 21 de julio de 2008].
- Björneborn, L., & Ingwersen, P. (2004). Toward a basic framework for webometrics. *Journal of the American Society for Information Science and Technology*, 55(14), 1216-1227.
- Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30, 1-7.
- Chau, M., Shiu, B., Chan, I., & Chen, H. (2007). Redips: Backlink search and analysis on the Web for business intelligence analysis. *Journal of the American Society for Information Science and Technology*, 58(3), 351-365.
- Chen, D., Hu, N., & Liu, L., (2007). Corporate blogging and firm performance: An empirical study. *Proceedings of the International Conference on Wireless Communications, Networking and Mobile Computing (WiCom 2007)* (pp. 6158-6161). Washington, DC: IEEE.
- Choi, H., & Varian, H. (2009). *Predicting the Present with Google Trends*. Disponible en: http://google.com/googleblogs/pdfs/google_predicting_the_present.pdf [consultado el 10 de mayo de 2009].
- Choi, H., & Varian, H. (2012). Predicting the Present with Google Trends. *Economic Record*, 88, 2-9.
- Chu, H., He, S., & Thelwall, M. (2002). Library and information science schools in Canada and USA: A Webometric perspective. *Journal of Education for Library and Information Science*, 43, 110-125.
- Da, Z., Engelberg, J., & Gao, P. (2011). In Search of Attention. *Journal of Finance*, 66(5), 1461-1499.
- Das, S.R., & Sisk, J. (2005). Financial communities. *Journal of Portfolio Management*, 31(4), 112-123.
- Faba, C., Guerrero, V., & Moya, F. (2004). *Fundamentos y técnicas cibernéticas*. Mérida: Junta de Extremadura. Disponible en: <http://www.juntaex.es/consejerias/economia-comercio-innovacion/dg-telecomunicaciones-sociedad-informacion/Publicaciones/common/tecnicasciberneticas.pdf> [consultado el 21 de julio de 2008].
- Garfield, E. (1979). *Citation indexing: Its theory and applications in science, technology and the humanities*. New York: Wiley, Interscience.
- Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457, 1012-1014 (February 19, 2009).
- Google blog (2008). *We knew the web was big...* Disponible en: <http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html> [consultado el 14 de enero de 2009].
- Holmberg, K. (2009). *Webometric Network Analysis. Mapping Cooperation and Geopolitical Connections between Local Government Administration on the Web*. Doctoral dissertation. Åbo: Åbo Akademi University Press. Disponible en: <http://kimholmberg.fi/phd/ThesisFinal.pdf> [consultado el 30 de noviembre de 2009].
- Huang, H., & Penna, N.D. (2009). *Constructing Consumer Sentiment Index for U.S. Using Google Searches*. Technical report, University of Alberta. Disponible en: http://econpapers.repec.org/paper/risalbaec/2009_5f026.htm [consultado el 17 de enero de 2013].
- Ingwersen, P. (1998). The calculation of Web Impact factors. *Journal of Documentation*, 54(2), 236-243.
- Kholodilin, K. A., Podstawski, M., & Siliverstovs, B. (2010). *Do Google searches help in nowcasting private consumption?: a real-time evidence for the US*. KOF Working Papers. Disponible en: <http://www.econbiz.de/en/search/detailed-view/doc/all/do-google-searches-help-in-nowcasting-private-consumption-a-real-time-evidence-for-the-us-kholodilin-konstantin/10008729133> [consultado el 21 de enero de 2013].
- Kirkpatrick, D., Roth, D., & Ryan, O. (2005) (10 de enero). Why there's no escaping the blog: Freewheeling bloggers can boost your product—Or destroy it. *Fortune*, 44.
- Lambert, F. (2008). *Rewriting the "Rules" of Online Networked Community Information Services: A Case Study of the mycommunityinfo.ca Model*. Canada : PhD dissertation, University of Western Ontario.
- Li, X. (2003). A review of the development and application of the Web impact factor. *Online Information Review*, 27(6), 407-417.
- Liu, B., Ma, Y., & Yu, P.S. (2001). Discovering unexpected information from your competitors' Web sites. *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 144-153). NewYork: ACM Press.
- Machill, M., Beiler, M., & Zenker, M. (2008). Search-engine research: a European-American overview and systematization of an interdisciplinary and international research field. *Media, Culture & Society*, 30(5), 591-608.
- O'Reilly, T. (2005). *What is Web 2.0. Design Patterns and Business Models for the Next Generation of Software*. Disponible en: <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html> [consultado el 17 de enero de 2009].
- Preis, T., Reith, D., & Eugene, H. (2010). Complex Dynamics of Our Economic Life on Different Scales: Insights from Search Engine Query Data. *Philosophical Transactions of the Royal Society of London*, 368, 5707-5719. Disponible en: <http://rsta.royalsocietypublishing.org/content/368/1933/5707> [consultado el 17 de enero de 2013].
- Radinsky, K., Davidovich, S., & Markovitch, S. (2009). Predicting the News of Tomorrow Using Patterns in Web Search Queries. *Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence (WIO8)*. Disponible en: <http://portal.acm.org/citation.cfm?id=1487070> [consultado el 23 de enero de 2013].
- Ravid, G., Bar, J., Rafaeli, S., & Baruchson, S. (2007). Popularity and findability through log analysis of search terms and queries: The case of a multilingual public service Web site. *Journal of Information Science*, 33(5), 567-583.
- Reid, E. (2003). Using Web link analysis to detect and analyze hidden Web communities, en: D. Vriens (Ed.). *Information and communications technology for competitive intelligence*, Hilliard, OH: Ideal Group, 57-84.
- Romero, E. (2010). *Una perspectiva webmétrica del estudio de empresas. Aplicación al estudio de variables financieras en empresas con presencia en la Web*. Tesis doctoral. Universidad de Granada.

- Smith, A., & Thelwall, M. (2002). Web Impact Factors for Australasian universities. *Scientometrics*, 54, 363-380.
- Tang, R., & Thelwall, M. (2003). U.S. academic departmental web-site interlinking in the United States: Disciplinary differences. *Library & Information Science Research*, 25(4), 437-458.
- Tang, R., & Thelwall, M. (2004). Patterns of national and international Web inlinks to US academic departments: An analysis of disciplinary variations. *Scientometrics*, 60(3), 475-485.
- Thelwall, M. (2000). Web impact factors and search engine coverage. *Journal of Documentation*, 56(2), 185-189.
- Thelwall, M. (2001). Extracting macroscopic information from Web links. *Journal of the American Society for Information Science and Technology*, 52(13), 1157-1168.
- Thelwall, M. (2002). Evidence for the existence of geographic trends in university Web site interlinking. *Journal of Documentation*, 58(5), 563-574.
- Thelwall, M. (2008). Bibliometrics to webometrics. *Journal of Information Science*, 34(4), 605-621.
- Thelwall, M. (2009). *Introduction to Webometrics. Quantitative Web Research for the Social Sciences*. Morgan & Claypool.
- Thelwall, M., Vaughan, L., & Björneborn, L. (2005). Webometrics, en: B. Cronin, (Ed.) *Annual review of information science and technology*, Medford, NJ: Information Today, 81-135.
- Thelwall, M., Vaughan, L., Cothey, V., Li, X., & Smith, A.G. (2003). Which academic subjects have most online impact? A pilot study and a new classification process. *Online Information Review*, 27(5), 333-343.
- Vaughan, L. (2004a). Web hyperlinks reflect business performance—A study of US and Chinese IT companies. *Canadian Journal of Information and Library Science*, 28(1), 17-31.
- Vaughan, L. (2004b). Exploring website features for business information. *Scientometrics*, 61(3), 467-477.
- Vaughan, L. (2004c). New measurements for search engine evaluation proposed and tested. *Information Processing & Management*, 40(4), 677-691.
- Vaughan, L. (2009). *Statistical Methods for the Information Professional*. Information Today. Medford, New Jersey.
- Vaughan, L., & Hysen, K. (2002). Relationship between links to journal web sites and Impact Factors. *ASLIB Proceedings: New Information Perspectives*, 54, 356-361.
- Vaughan, L., & Romero, E. (2010). Web hyperlink patterns and the financial variables of the global banking industry. *Journal of Information Science*, 36(4), 530-541.
- Vaughan, L., & Thelwall, M. (2003). 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*, 54, 29-38.
- Vaughan, L., & Wu, G.Z. (2004). Links to commercial websites as a source of business information. *Scientometrics*, 60(3), 487-496.
- Vaughan, L., & You, J. (2009). Keyword enhanced Web structure mining for business intelligence. *Lecture Notes in Computer Science*, 4879, 161-168.
- Vaughan, L., Tang, J., & Du, J. (2010). Constructing Business Profiles Based on Keyword Patterns on Web Sites. *Journal of American Society for Information Science*, 61(6), 1120-1129.
- Vosen, S., & Schmidt, T. (2011). Forecasting private consumption: survey-based indicators vs. Google trends. *Journal of Forecasting*, 30(6), 565-578.
- Vreeland, R.C. (2000). Law libraries in hyperspace: A citation analysis of World Wide Web sites. *Law Library Journal*, 92(1), 9-25.
- Wilkinson, D., Harries, G., Thelwall, M., & Price, E. (2003). Motivations for academic web site interlinking: evidence for the web as a novel source of information on informal scholarly communication. *Journal of Information Science*, 29(1), 59-66.
- Wormell, I. (2001). Informetrics and Webometrics for measuring impact, visibility, and connectivity in science, politics, and business. *Competitive Intelligence Review*, 12(1), 12-23.
- Wu, L., & Brynjolfsson, E. (2010). The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales. Technical report, MIT. Disponible en: http://www.nber.org/confer/2009/PRf09/Wu_Brynjolfsson.pdf [consultado el 17 de enero de 2013].

