

Cibermetría. Estado de la cuestión

Natalia Arroyo, José Luis Ortega, Víctor Pareja, José Antonio Prieto, Isidro Aguillo

Laboratorio de Internet

C/ Joaquín Costa, 22 despacho 313
28002 Madrid (España)

Tel.: +34 915635482 (ext. 313)

narroyo@fundaciongsr.es, {jortega, vmpareja, joseaprieto, isidro}@cindoc.csic.es

Resumen

De la aplicación de teorías y técnicas procedentes de la Bibliometría, Cienciometría e Informetría al estudio de Internet, surgen, a finales de los noventa la Cibermetría y la Webmetría, con el objeto de describir cuantitativamente los contenidos y procesos de comunicación que se producen en el ciberespacio y en el Web respectivamente. En el presente trabajo se hará un recorrido por la literatura que esta joven disciplina ha producido, desde sus inicios hasta la actualidad, con el fin de identificar la trayectoria que ha seguido hasta el momento y los frentes de investigación que permanecen abiertos. En la vertebración de este estudio se referirán los siguientes puntos: definición, conceptualización y ámbito de la disciplina, técnicas que emplea, contribuciones de otras disciplinas, y problemática que presenta, para finalizar con un estado de la situación en nuestro país.

Palabras clave: cibermetría, webmetría, revisión, líneas de investigación, métodos

1. Definición y contexto de la Cibermetría

El origen de la Cibermetría puede situarse a mediados de los noventa, cuando, marcada por un fuerte carácter descriptivo, sus principales objetivos consistían en el estudio de aspectos como la evolución del tamaño del Web y la descripción de los primeros motores de búsqueda. Es el reconocimiento de las dos principales características del Web, su cobertura global y su naturaleza hipertextual, lo que propicia la publicación de una serie de trabajos en los que se aplican los principios de la Bibliometría e Informetría al estudio del Web, así como la posterior aparición de la revista electrónica *Cybermetrics* en 1997, presentada en el primero de una serie de seminarios sobre la diseminación de resultados del análisis cuantitativo de Internet (Aguillo, 1997), que se celebran con carácter bianual en el marco de las conferencias de la International Society for Scientometrics and Informetrics (ISSI).

En sus inicios fueron propuestos varios términos para designar a la nueva disciplina, como señala Björneborn (2004): netometrics (Bossy, 1995), webometry (Abraham, 1998), internetometrics (Almind e Ingwesen, 1996), web bibliometry (Chakrabarti, et al., 2002), aunque finalmente se adoptaron dos, que, sin llegar a serlo, a menudo se emplean como sinónimos: cybermetrics, que toma su nombre de la revista electrónica, y webometrics (Almind e Ingwesen, 1997). Para su traducción al español ambos fueron adaptados literalmente del inglés, dando así lugar a las expresiones Cibermetría y Webmetría.

El matiz que los distingue es su ámbito de actuación, tal y como queda reflejado en las definiciones que Björneborn (2004) propone. Según él, la Cibermetría puede ser entendida como “el estudio de los aspectos cuantitativos de la construcción y uso de los recursos de información, estructuras y tecnologías en Internet, desde perspectivas bibliométricas e informétricas”, mientras que la Webmetría puede definirse exactamente de la misma forma, con la salvedad de que se refiere sólo a una parte de Internet, el Web.

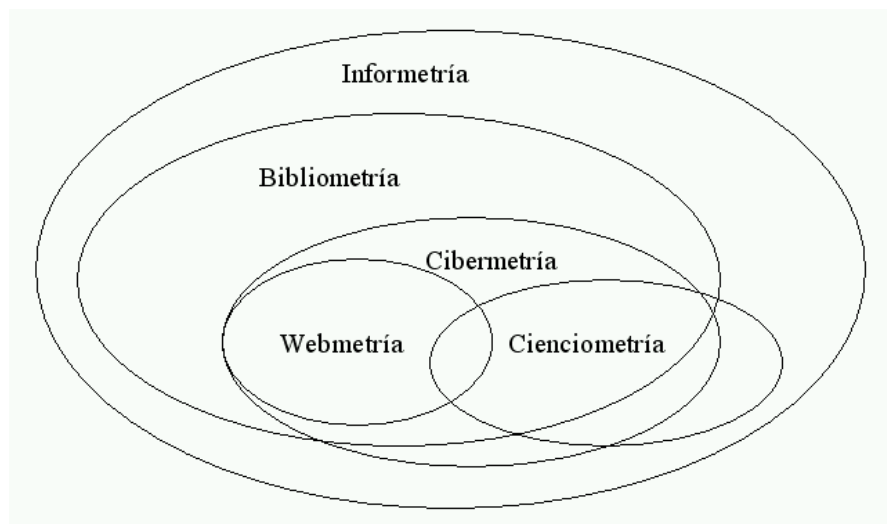


Figura 1. Relación entre Bibliometría, Cienciometría, Informetría, Webmetría y Cibermetría. Adaptado de Björneborn (2002).

La relación de la cibermetría con otras ciencias afines —Bibliometría, Cienciometría e Informetría— es de solapamiento en algunas áreas, ya que comparten un mismo enfoque y un mismo ámbito, enmarcado por la Informetría, y que es por lo tanto el de los estudios cuantitativos sobre la información (figura 1).

2. Técnicas empleadas

La aplicación de técnicas bibliométricas a Internet, sobre las que se fundamenta la Cibermetría, se basa en el principio según el cual se puede considerar cierta equivalencia entre una cita bibliográfica y un hipervínculo enlazando a una página web. Desde esta perspectiva, y por analogía con las citas bibliográficas, se acuñó el término “sitations” (McKiernan, 1996; Aguillo,

1996; Rousseau, 1997), o “sitas” en español, para designar a los enlaces entre diferentes sitios web.

Otros trabajos que aparecieron en esta misma línea y de los que se puede decir que sentaron las bases de la disciplina son los de Bray (1996), Larson (1996), Abraham (1996; 1998), o Almind e Ingwersen (1997).

Un punto clave que algunos autores comenzaron poco después a cuestionarse, es si la aplicación de las leyes clásicas de la Informetría a Internet es válida o no (Boudourides, Sigrist y Alevizos, 1999; Egghe, 2000), a la que este último respondió que esto es posible siempre que se produzcan dos condiciones: que haya una fuente (revistas, autores, artículos...) y un ítem a medir (artículos, citas o referencias...), tal y como sucede en Cienciometría, y esto también se da en el caso de las revistas electrónicas, o cuando existe una página web que actúa como fuente y al menos un hipervínculo en ella, lo que sería un ítem, tal y como expuso Rousseau (1997).

En ese mismo trabajo se demuestra que la Ley de Lotka se da en la distribución de sitas a un sitio web y de dominios de primer nivel por sitio web. La misma ley fue observada por Boudourides y Antipas (2002) en la distribución del número de páginas por sitios web.

Sobre otras leyes clásicas de la Cienciometría incidieron Bar-Ilan (1997), quien aplica la Ley de Bradford a los grupos de noticias en un estudio sobre la enfermedad de las vacas locas, empleando para ello un conjunto de datos extraídos mediante AltaVista, y llegando a la conclusión de que la distribución de Bradford se puede aplicar a estos grupos de noticias e identificar núcleos; o Faba, Guerrero y Moya (2003) en un estudio sobre el ajuste de los datos de sitas a esta misma distribución, entre otros. También otras leyes exponenciales han sido aplicadas, a enlaces entrantes y salientes (Albert, Jeong y Barabasi, 1999), número de páginas por sitio (Adamic y Huberman, 2000). Además, la Ley de Zipf ha sido extrapolada a la distribución de enlaces entrantes (Broder et al., 2000), la red que forman correos electrónicos enviados y recibidos, la conectividad de los routers en Internet (Adamic y Huberman, 2002), o ya más recientemente a los correos enviados a los grupos de noticias (Kot, Silverman y Berg, 2003).

En lo que a las leyes de crecimiento exponencial propuestas por Price (1963) sobre la ciencia, se cumplen, tal y como Egghe (2000) identificó, también en aspectos de la Red como el número de servidores o el número de hosts, según datos proporcionados por Zakon (2003) y el Internet Software Consortium (2000).

Otras técnicas procedentes de la Cienciometría que también se han aplicado al estudio del Web son el análisis de citas, cocitas y de co-ocurrencia de palabras. El primero de ellos, el

análisis de enlaces web —o citas— como si de citas bibliográficas se tratara ha encontrado oposición entre algunos autores (Egghe, 2000; Van Raan, 2001), que consideran que la motivación es distinta y no sería correcta la comparación, mientras que otros opinan que se ha encontrado un equivalente a las bases de datos tradicionales (Rodríguez i Gairín, 1997; Almind e Ingwersen, 1997).

El análisis de co-citas, basado en el principio según el cual a mayor número de veces en que dos documentos son citados conjuntamente, mayor posibilidad existe de que su contenido esté relacionado, fue primeramente adaptado al Web por Larson (1996) al crear un mapa de co-citación en el área de ciencias de la Tierra mediante la técnica de escalamiento multidimensional (MDS). Posteriormente otros autores lo han empleado en este ámbito con diferentes fines: Dean y Henzinger (1999) para la búsqueda de páginas relacionadas en el Web; Boudourides, Sigrist y Alevizos (1999) para el estudio de las relaciones entre los participantes en el proyecto SOEIS; Kumar et al. (1999), en la identificación de comunidades también en el Web; y ya más recientemente Thelwall y Wilkinson (2004) en la identificación de sitios web que guardan cierta similitud. En la actualidad Katz desarrolla un estudio sobre indicadores de co-enlaces, entre los que distingue hosts, URLs y regiones geográficas co-enlazados (Katz, 2004).

El análisis de co-ocurrencia de palabras, ya por último, parte de la premisa de que los términos que aparecen en un texto representan el contenido del mismo, siendo así posible medir el número de veces en que aparecen conjuntamente en diferentes documentos con el fin de generar redes conceptuales de los diferentes campos del saber. También ha sido aplicado al campo de la Cibermetría en estudios como los de Ross y Wolfram (2000), quienes estudiaron pares de términos enviados a un motor de búsqueda, Excite; o Leydesdorff y Curran (2000), cuyo objetivo era conocer la relación entre los componentes de la “triple hélice” mediante la co-ocurrencia de los términos “university”, “industry” y “government” en páginas de diferentes dominios.

Aunque el Web es el servicio más estudiado dentro de la Cibermetría, es de justicia mencionar también otros trabajos realizados sobre las redes de comunicación en otros ámbitos, como los grupos de noticias —recordemos de nuevo los trabajos de Bar-Ilan (1997) y Faba, Guerrero y Moya (2003) — o el e-mail (Drineas et al., 2004; Zelman y Leydesdorff, 2000).

3. Indicadores web

Otro de los frentes abiertos en el campo de la Cibermetría es la construcción de indicadores. De ellos el que más líneas ha ocupado, quizás por su indudable importancia para la Bibliometría, es el factor de impacto web (o Web Impact Factor en la literatura anglosajona), que no es más que una aplicación a esta disciplina del factor de impacto de Garfield. Aunque el concepto de impacto de la información en el Web fue en un principio introducido por Rodríguez

i Gairín (1997), fue Ingwersen (1998) quien finalmente lo definió como “la suma lógica del número de enlaces externos y auto-enlaces que apuntan a un determinado país o sitio web, dividido por el número de páginas en ese país o sitio web en un momento determinado. El numerador consiste, por lo tanto, en el número de enlaces a páginas, no de enlaces”.

El mismo Ingwersen habla además de un factor de impacto web externo, que calcula extrayendo las auto-sitas del recuento total, evitando así uno de los problemas del factor de impacto. Varios autores (Egghe, 2000; Li, 2003; Smith, 1999; Smith y Thelwall, 2002; Thelwall, 2002b; Vaughan y Hysen, 2002) han continuado esta línea con diversas aportaciones de indudable interés pero cuya extensión sobrepasaría los límites de este trabajo. Sin embargo, sí resulta digna de mención la correlación hallada entre los índices del Research Assessment Exercise (RAE) y los factores de impacto web de varios departamentos universitarios en el Reino Unido (Li et al., 2003).

La relevancia de los indicadores web ha sido puesta de manifiesto con el ya finalizado proyecto del V Programa Marco de Investigación y Desarrollo de la Comisión Europea EICSTES (European Indicators, Cyberspace and the Science-Technology-Economy System), en el que fueron empleados en el análisis del sistema europeo de ciencia, tecnología y economía en Internet para establecer las relaciones entre el sector de I+D y los actores de la nueva economía. Y, es más, parece consolidarse con WISER (Web Indicators for Science, Technology and Innovation Research), también parte del V Programa Marco, que tiene por objetivo explorar las posibilidades y problemas en el desarrollo de una nueva generación de indicadores de ciencia y tecnología basados en el Web. En él se da la siguiente definición de indicador web: “una medida de relevancia en la toma de decisiones que cuantifica aspectos de la creación, difusión y aplicación de la ciencia y la tecnología en la medida en que están representadas en Internet o el World Wide Web”.

4. El enfoque de otras disciplinas

Además de los enfoques próximos a la Informetría, Cienciometría y Bibliometría de los que se ha hablado hasta ahora, es necesario también tener en cuenta otras aportaciones de diversos campos más cercanos al entorno de las ciencias, como la informática, las física, las matemáticas o las ciencias sociales. Quizás la que más implicaciones ha supuesto, más bien por su valiosa aportación a los algoritmos de extracción de datos, búsqueda e identificación de comunidades, y como fenómeno sociológico, en palabras de Broder et al. (2000), es la aplicación de la teoría de grafos al análisis de la estructura del Web (aunque también a otros procesos de comunicación en Internet). Este es considerado como si de un grafo dirigido se tratara, en el cual las páginas web serían nodos o vértices y los hipervínculos que las relacionan los arcos o aristas. Uno de los primeros autores en explotar esta perspectiva fue Abraham (1996; Abraham y Foresta, 1996), cuyas estrategias de medida y visualización del Web están basadas en la teoría de sistemas dinámicos complejos.

Sobre la estructura del Web han sido extraídas interesantes conclusiones acerca de sus características: Broder et al. (2000) lo dibujan como una especie de corbata de lazo (bow-tie) formado por un núcleo de páginas altamente interconectadas, unos lazos —con enlaces dirigidos al núcleo uno y que salen de él en el otro—, un delgado tubo que conecta ambos lazos, y algunos hilos sin procedencia definida. Estructuras similares han sido halladas más recientemente tanto en el Web chileno (Baeza-Yates, 2003) y el Web académico del Reino Unido (Björneborn, 2004), este último con forma de corona.

El diámetro web, entendido como el número mínimo de enlaces por los que hay que pasar llegar desde un documento a otro, es de 19, según Albert, Jeong y Barabasi (1999), un número sorprendentemente bajo que demuestra la alta conectividad que caracteriza a la Red. Kleinberg y Lawrence (2001) identificaron dos tipos de páginas web: intercambiadores y autoridades; los primeros son aquellos que apuntan a muchas autoridades, y éstos, a su vez, los enlazados por muchos intercambiadores. Los patrones de ambos, tal y como estos autores afirman, pueden ser empleados para determinar comunidades de páginas relacionadas por un mismo tema mediante técnicas de clusterización.

Fueron Watts y Strogatz (1998) quienes manifestaron que el Web cumple la característica de algunos grafos conocida como small-world, por analogía con el fenómeno así llamado (también conocido popularmente como “seis grados de separación”) que consiste en un estado intermedio entre la regularidad y la aleatoriedad en la distribución de los arcos de un grafo, lo que da lugar a un alto grado de agrupación o clusterización. Este trabajo abrió una línea de investigación continuada entre otros por Adamic (1999), Barabasi (2003) y Björneborn (2002; 2004).

Otras aproximaciones, algunas desde las teorías de redes, tales como el análisis de redes sociales (Social Network Analysis), —basada en que las relaciones o interacciones entre actores son los factores clave que sostienen y definen una red, como puede ser Internet (Wellman et al., 1996)—, Redes Neuronales Artificiales (RNA) —en la identificación de clusters y la creación de mapas (Kohonen, 1995; Polanco, François y Lamirel, 2001)—, teoría de redes complejas (Scharnhorst, 2004); y otras como las teorías del caos —el análisis caótico ha sido aplicado a las series temporales de datos obtenidos de Internet, tales como datos de tráfico de red o de datos de caché (Kugiumtzis y Boudourides, 1998)— y fractales —(Abraham, 1996; Abraham y Foresta, 1996; Egghe 1997), el último de los cuales hace un estudio de la dimensión fractal de los sistemas de hipertexto—, también han sido aplicadas al estudio del Web.

5. Problemática de la Cibermetría

Internet ha sido caracterizado en numerosas ocasiones como un medio dinámico y cambiante, rasgo que ha sido objeto de estudio en numerosas ocasiones. Ha sido Koehler, investigador de la Universidad de Oklahoma, quien más ha profundizado en este aspecto; él establece dos tipos de cambios en los documentos web: persistencia —que define como la existencia o desaparición de páginas y sitios web— e intermitencia —variante de la persistencia definida como la desaparición y posterior reaparición de documentos Web, que considera se dan en un momento concreto sobre alrededor de un 5% de las páginas web (Koehler, 1999b). Posteriormente este autor ha venido desarrollando, en esta misma línea, otros indicadores del cambio en este ámbito, tales como consistencia y permanencia (Koehler, 1999a, 2002).

Las herramientas disponibles en la actualidad y la propia naturaleza de Internet hacen imposible determinar exactamente cuál es su tamaño. En lo que al tamaño físico se refiere existe información institucional muy precisa sobre la evolución de las estructuras en Internet, como el número de hosts (Internet Systems Consortium, Inc., RIPE Network Coordination Centre), o de dominios (Country Registry Resources), pero sobre el número exacto de páginas sólo pueden realizarse estimaciones. Lawrence y Giles (1999) consideraron que en febrero de ese mismo año había unos 800 millones de páginas web distribuidos en un total de 3 millones de servidores, de los cuales tan sólo un 6% tenía contenidos científicos o educativos. En julio de 2000 Moore y Murray hablan de un total de 2.200 millones de páginas web, con un incremento mensual de 7 millones de páginas. Hasta el momento la única posibilidad es extraer datos sectoriales (por dominios o sub-dominios, lenguas, etc.) interrogando a los motores de búsqueda (Zakon, 2003).

Sin embargo, una de las limitaciones de estos, es su imposibilidad para cubrir todo el Web, dando así lugar a lo que se conoce como Web invisible —Internet invisible para referirse a todo el ciberespacio—, también llamada “deep web” (Bergman, 2001) o “dark matter” (Bailey, Craswell y Hawking, 1999), por no ser indexable para los motores de búsqueda, y cuyo tamaño ha sido estimado entre unas 400 y 550 veces mayor que el del Web visible, como si este último sólo fuera la punta de un iceberg.

Sherman y Price, por su parte, lo definen de la siguiente manera: “páginas textuales, ficheros u otra información fidedigna, a menudo de alta calidad, disponible en el World Wide Web, que los motores de búsqueda no pueden, debido a limitaciones técnicas, o no podrán, por elección deliberada, añadir a sus índices de páginas web” (Sherman y Price, 2001). Dichos autores distinguen cuatro tipos de invisibilidad: Web opaco, Web privado, Web propietario y el Web realmente invisible.

Una de las cuestiones que más atención han atraído son los métodos de extracción de datos, muy especialmente los motores de búsqueda, que han sido, hasta el momento, las herramientas más empleadas por los cibermétricos en la extracción de datos del Web.

Numerosos estudios han venido investigando sus características —tamaño y cobertura (Bharat y Broder, 1998; Lawrence y Giles, 1998; Lawrence y Giles, 1999; Notess, 2003; Vaughan y Thelwall, 2004), estabilidad (Bar-Ilan, 1998), solapamiento (Notess, 2002), entre otros— y siguiendo su evolución en el tiempo (Thelwall, 2001a).

En un principio AltaVista constituyó el centro de atención por sus ventajas competitivas con respecto de los demás, pero desde su aparición ha sido Google quien ha atraído la mayor parte de las miradas en un principio por la novedad de su PageRank (Brin y Page, 1998), que se fundamenta en la idea, muy afín a los postulados bibliométricos, de que el número de enlaces que apuntan a una página determinará su posición —de ahí el nombre de “posicionamiento web” que recibe el área dedicada al estudio de la visibilidad de los sitios web en los distintos motores de búsqueda— en los resultados de una búsqueda, ocupando los primeros puestos si el número es mayor. Sin embargo, las limitaciones de este tipo de herramientas son muchas: no cubren todo el Web, ni siquiera la parte que es técnicamente indexable, los algoritmos que emplean están protegidos por el secreto comercial, por lo que su funcionamiento es un misterio; la validez de los datos que arrojan es cuestionable dadas las grandes diferencias encontradas incluso en datos recogidos en intervalos de tiempo mínimos (Bar-Ilan, 1998; Rousseau, 1999; Snyder y Rosenbaum, 1999); y además existen sesgos nacionales en la cobertura de los sitios, tal y como se ha demostrado recientemente (Vaughan y Thelwall, 2004).

Todo esto, unido a la búsqueda de métodos alternativos, ha llevado a algunos grupos de investigación a desarrollar sus propias herramientas, que permiten, como se demostrará más adelante, un mayor control sobre los resultados. Es el caso de los *crawlers* SocSciBot, diseñado por el *Statistical Cybermetrics Research Group* de la Universidad de Wolverhampton (Thelwall, 2001b), y Webcount, uno de los resultados del proyecto EICSTES (Adams y Gilbert, 2003).

Otra alternativa, propuesta por Aguillo (1998b; 2000b), es el uso de herramientas de segunda generación, en concreto programas comerciales de bajo coste (shareware) como mapeadores o verificadores de enlaces, que han sido empleados para la extracción de datos de sedes web en varios proyectos de investigación. La diferencia de estas dos últimas soluciones reside en la perspectiva o nivel del análisis: si bien los motores de búsqueda son más apropiados para extraer información a un nivel macro (grandes cantidades de información), los *crawlers* académicos y comerciales se basan en unidades de análisis menores, como sitios o sedes web (nivel micro), lo que hace más costoso su trabajo para grandes volúmenes de información.

Otro punto importante, en el que nadie parece ponerse de acuerdo debido a la dificultad que entraña el ajustar la diversidad de criterios adoptados por los webmasters a unos patrones determinados, y muy ligado al tema anterior, es el referente a las unidades de medida. Teniendo en cuenta que la mayor parte de los estudios realizados se basan en los motores de

búsqueda, y estos, a su vez, indexan páginas web, ha sido esta la unidad más empleada hasta el momento. Sin embargo otros han preferido emplear el concepto, un tanto confuso por su falta de conceptualización, de sitio web.

Por ello, y desde un punto de vista más documental, Aguillo (1998a) propuso el concepto de sede web definido por Arroyo y Pareja (2003) como “página web, o conjunto de páginas web ligadas jerárquicamente a una página principal, identificable por una URL y que forma una unidad documental reconocible e independiente de otras bien por su temática, bien por su autoría, o por su representatividad institucional. Teniendo en cuenta este último aspecto se reconocerían tres tipos de sedes web: institucionales, temáticas y personales”. Si bien es verdad que este planteamiento plantea una gran dificultad al trabajar con grandes cantidades de información, ya que no es posible automatizar el proceso de identificación, siendo por ello más adecuado para estudios a nivel micro, también lo es que permite ordenar el conocimiento en el web desde un punto de vista documental, y que admite la existencia de varios niveles jerárquicos. Por ello ha sido aplicado en varios estudios (Pareja, González y Aguillo, 1999; Aguillo, 2000; Arroyo, Pareja y Aguillo, 2003).

Thelwall, por su parte, propone el empleo de lo que él llama “modelos alternativos basados en el documento” o ADMs —Alternative Document Models—, que define como “una unidad de trabajo con un tema identificable consistente, producida por uno o varios autores. Su acceso puede estar restringido a algunas partes o al documento completo, que es recuperable en el Web mediante un navegador actualizado” (Thelwall, 2002b), entre los que incluye cuatro: página web individual, directorio, nombre de dominio y sitio. La ventaja de esta propuesta reside en que este tipo de unidades pueden ser identificadas y extraídas automáticamente, pero, por otra parte, cabe decir que el punto de vista desde el este modelo ha sido creado es más informático que documental, y como consecuencia las cuatro unidades definidas vienen más bien determinadas por la propia arquitectura del Web. Pero quizás la mayor objeción es que resulta un tanto cuestionable la identificación de ciertas unidades con documentos, en concreto página web y directorio, especialmente desde un punto de vista de los contenidos.

6. La Cibermetría en España

En España la Cibermetría goza en la actualidad de un sorprendente desarrollo gracias a la amplia labor de difusión realizada, tal y como puede constatarse tras la aparición de varios artículos publicados en revistas nacionales e internacionales —la mayor parte de los cuales ya han sido citados anteriormente—, dos tesis doctorales (Alonso, 2002; Faba, 2002), que además han dado lugar a sendas monografías recientemente publicadas (Alonso, Figuerola y Zazo, 2004; Faba, Guerrero y Moya, 2004), y la contribución en varios proyectos de investigación nacionales y europeos (ICYTNET, EICSTES, WISER), así como su enseñanza en algunas universidades, como es el caso de la Universidad de Extremadura.

Bibliografía citada

- Abraham, R. H. (1996). "Webometry: measuring the complexity of the World Wide Web". *World Futures*, 50, 785-791. <<http://www.ralph-abraham.org/articles/MS%2385.Web1/>>. [Consulta: 28/10/2004].
- Abraham, R. H. (1998). "Webometry: measuring the synergy of the World Wide Web". *Biosystems*. 46(1-2), 209-212.
- Abraham, R. H.; Foresta, D. (1996). "Webometry: chronotopography of the World Wide Web". <<http://www.ralph-abraham.org/articles/MS%2389.Web3/>>. [Consulta: 28/10/2004].
- Adamic, L. A. (1999). "The small world Web". *Proceedings of ECDL'99. Lecture Notes in Computer Science*. 1696, 443-452.
- Adamic, L. A.; Huberman, B. A. (2000). "Power-law distribution of the World Wide Web". *Science*. 287, 2115-2116.
- Adamic, L. A.; Huberman, B. A. (2002). "Zipf's law and the Internet". *Glottometrics*. 3, 143-150.
- Adams, K.; Gilbert, N. (2003). "Indicators of intermediaries' role and development". Deliverable 6.2. Proyecto EICSTES.
- Aguillo, I. (1996). "A Preliminary Approach to Citation Phenomena in the World Wide Web". *EASST / 4S Conference, Bielefeld*, 10-13 octubre 1996.
- Aguillo, I. (1998a). "Hacia un concepto documental de sede web". *El Profesional de la Información*. 7(1-2), 45-46.
- Aguillo, I. (1998b). "Herramientas de segunda generación". *Anuario SOCADI*, 85-112.
- Aguillo, I. (2000). "Indicadores: hacia una evaluación no objetiva (cuantitativa) de sedes web". *Jornadas Españolas de Documentación. FESABID*. 7, 233-248.
- Albert, R.; Jeong, H.; Barabasi, A. L. (1999). "Diameter of the World Wide Web". *Nature*. 401(6749), 130-131.
- Almind, T.; Ingwersen, P. (1996). "Informetric analysis on the World Wide Web: a methodological approach to "internetometrics"". Centre for Informetric Studies, Royal School of Library and Information Science (CIS Report 2).
- Almind, T. C.; Ingwersen, P. (1997). "Informetric analyses on the World Wide Web: metodological aproaches to "Webometrics"". *Journal of Documentation*, 53(4), 404-426.
- Alonso, J. L. (2002). "Cibermetría: análisis de los dominios web españoles". Salamanca: Universidad de Salamanca. Tesis doctoral.
- Alonso, J. L.; Figuerola, C. G.; Zazo, A. F. (2004). "Cibermetría: nuevas técnicas de estudio aplicables al Web". Gijón: TREA.
- Arroyo, N.; Pareja, V. M. (2003). "Metodología para la obtención de datos con fines cibernéticos". *III Taller de Indicadores Bibliométricos*; Madrid, 3-5 marzo. <<http://internetlab.cindoc.csic.es/varios/Metodolog%EDa%20datos%20ciberm%EA9tricos.pdf>>. [Consulta: 28/10/2004].

- Arroyo, N.; Pareja, V. M.; Aguillo, I. (2003). "Description of web data in D3.1". Proyecto EICSTES. Deliverable 3.2.
- Baeza-Yates, R. (2003). "Evolución de la Web Chilena 2001-2002" <<http://www.ciw.cl/recursos/estudio2002/>>. [Consulta: 28/10/2004].
- Bailey, P.; Craswell, N.; Hawking, D. (2000). "Chart of darkness: mapping a large intranet". CSIRO CMIS Technical Report.
- Bar-Ilan, J. (1997). "The "mad cow disease", usenet groups and bibliometric laws". *Scientometrics*. 39(1), 1997.
- Bar-Ilan, J. (1998). "Search engine results over time: a case study on search engine stability". *Cybermetrics*. 2/3 (1). <<http://www.cindoc.csic.es/cybermetrics/articles/v2i1p1.html>>. [Consulta: 28/10/2004].
- Barabasi, A. L. (2003). "Linked: how everything is connected to everything else and what it means for business, science, and everyday life". New York: Plume.
- Bharat, K.; Broder, A. (1998). "Measuring the Web". <<http://www.research.compaq.com/SRC/whatsnew/sem.html>>. [Consulta: 28/10/2004].
- Bergman, M. K. (2001). "White paper. The deep Web: surfacing hidden value". <<http://www.brightplanet.com/pdf/deepwebwhitepaper.pdf>>. [Consulta: 28/10/2004].
- Björneborn, L. (2002). "Small-world link structures on the Web". <<http://www.db.dk/lb/2002smallworld.pps>>. [Consulta: 28/10/2004].
- Björneborn, L. (2004). "Small-world link structures across an academic web space: a library and information science approach". Copenhagen: Department of Information Studies, Royal School of Library and Information Science. [Tesis doctoral] <<http://www.db.dk/lb/phd/phd-thesis.pdf>>. [Consulta: 28/10/2004].
- Bossy, M. (1995). "The last of the litter: Netometrics". *Solaris Information Communication*, 2, 245-250.
- Boudourides, M.; Antypas, G. (2002). "A simulation of the structure of the World Wide Web". *Sociological Research Online*, 7(1).
- Boudourides, M. A.; Sigrist, B.; Alevizos, P. D. (1999). "Webometrics and the self-organization of the European information society". Proyecto SOEIS, Rome meeting. <<http://hyperion.math.upatras.gr/webometrics/>>. [Consulta: 28/10/2004].
- Bray, T. (1996). "Measuring the Web". *Fifth International World Wide Web Conference*, Paris.
- Brin, S.; Page, L. (1998). "The anatomy of a large-scale hypertextual Web search engine". *Computers Networks and ISDN Systems*. 30(1-7), 107-117.
- Broder, A.; Kumar, R.; Maghoul, F.; Raghavan, P.; Rajagopalan, S.; Stata, R.; Tomkins A.; Wiener, J. (2000). "Graph structure in the Web". *Computer Networks*, 33(1-6), 309-320.
- Chakrabarti, S.; Joshi, M. L.; Punera, K.; Peenock, D. M. (2002). "The structure of broad topics on the Web". *WWW Conference* Honolulu, 7-11 mayo.

- Dean, J.; Henzinger, M. (1999). "Finding related pages in the World Wide Web". *Computers Networks and ISDN Systems*, 31, 389-401.
- Drineas, P.; Krishnamoorthy, S.; Sofka, M. D.; Yener, B. (2004). "Studying e-mail graphs for intelligence monitoring and analysis in the absence of semantic information". *Symposium on Intelligence and Security Informatics (ISI'04)*, junio.
- Egghe, L. (1997). "Fractal and informetric aspects of hypertext systems". *Scientometrics*, 40(3), 455-464.
- Egghe, L. (2000). "New informetric aspects of the Internet: some reflections, many problems". *Journal of Information Science*, 26(5), 329-335.
- EICSTES. <<http://www.eicstes.org>>. [Consulta: 28/10/2004].
- Faba, C. (2002). "Análisis cibernético de la información web: el caso de Extremadura en Internet". Granada: Universidad de Granada. [Tesis doctoral].
- Faba, C.; Guerrero, V. P.; Moya, F. (2003). "'Sitation' distributions and Bradford's law in a closed web space". *Journal of Documentation*, 59(5), 558-580.
- Faba, C.; Guerrero, V. P.; Moya, F. (2004). "Fundamentos y técnicas cibernéticas". Mérida: Junta de Extremadura.
- Ingwersen, P. (1998). "The calculation of Web Impact Factor". *Journal of Documentation*, 54(2).
- Internet Systems Consortium, Inc. <<http://www.isc.org/>>. [Consulta: 28/10/2004].
- Katz, S. (2004). "Co-Link Web Indicators of the European Research Area". Proyecto WISER, informe técnico.
- Kleinberg, J. M.; Lawrence, S. (2001). "The structure of the Web". *Science*, 294, 1849-1850.
- Koehler, W. (1999a). "An analysis of web page and web site constancy and permanence". *Journal of the American Society for Information Science and Technology*. 50(2), 162-180.
- Koehler, W. (1999b). "Digital libraries and World Wide Web sites and page persistence". *Information Research* 4(4).
- Koehler, W. (2002). "Web page change and persistence". *Journal of the American Society for Information Science and Technology*, 53(2), 162-171.
- Kohonen, T. (1995). "Self-organizing maps". Berlin: Springer.
- Kot, M.; Silverman, E.; Berg, C. A. (2003). "Zipf's law and the diversity of biology newsgroups". *Scientometrics*, 56(2), 247-257.
- Kugiumtzis, D.; Boudourides, M. A. (1998). "A chaotic analysis of Internet ping data: just a random number generator?". *SOEIS project Bielefeld conference*.
- Kumar, R.; Raghauam, P.; Rajagopalan, S.; Tomkins, A. (1999). "Crawling the Web for emerging cyber-communities". *Computer Networks*, 31(11), 1481-1493.
- Larson, R. R. (1996). "Bibliometrics of the World Wide Web: an exploratory analysis of the intellectual structure of the cyberspace". *Proceedings of the 59th Annual Meeting*; Baltimore, Maryland.
- Lawrence, S.; Giles, C. L. (1998). "Searching the World Wide Web". *Science*, 280, 98-100.

- Lawrence, S.; Giles, C. L. (1999). "Accesibility of information on the Web". *Nature*, 400, 107-109.
- Leydesdorff, L.; Curran, M. (2000). "Mapping university-industry-government relations on the Internet: the construction of indicators for a knowledge-based economy". *Cybermetrics*, 4(1).
- Li, X. (2003). "A review of the development and application of the Web Impact Factor". *Online Information Review*, 27(6), 407-417.
- Li, X.; Thelwall, M.; Musgrove, P.; Wilkinson, D. (2003). "The relationship between the WIFs or inlinks of computer science departments in UK and their RAE ratings or research productivities in 2001". *Scientometrics*, 57(2), 239-255.
- McKiernan, G. (1996). "CitedSites (sm): citation indexing of web resources". <<http://www.public.iastate.edu/~CYBERSTACKS/Cited.htm>>. [Consulta: 28/10/2004].
- Moore, A.; Murray, B. H. (2000). "Sizing the Web". Cyveillance, Inc. White Paper. <http://www.cyveillance.com/web/downloads/Sizing_the_Internet.pdf>. [Consulta: 28/10/2004].
- Notess, G. R. (2002). "Search Engines Statistics: Database Overlap". <<http://www.searchengineshowdown.com/stats/overlap.shtml>>
- Notess, G. R. (2003). "Search engine statistics: database total size estimates". <<http://www.notess.com/search/stats/sizeest.shtml>>. [Consulta: 28/10/2004].
- Pareja, V. M.; González, A.; Aguillo, I. (1999). "Ciencia y tecnología españolas en Internet: valoración a través de la presencia de organismos públicos españoles y de sus revistas electrónicas". *Arbor*, CLXII(639), 367-390.
- Polanco, X.; François, C.; Lamirel, J. C. (2001). "Using artificial neural networks for mapping of science and technology: a multi-self-organizing-maps approach". *Scientometrics*, 51(1), 267-292.
- Price, D. J. de S. (1963). "Little science, big science". New York: Columbia University Press.
- RIPE Network Coordination Centre. <<http://www.ripe.net/>>. [Consulta: 28/10/2004].
- Rodríguez i Gairín, J. M. (1997). "Valoración del impacto de la información en Internet: Altavista, el "Citation Index" de la Red". *Revista Española de Documentación Científica*, 20(2), 175-181.
- Ross, N. C. M.; Wolfram, D. (2000). "End user searching on the Internet: an analysis of term pair topics submitted to the Excite search engine". *Journal of the American Society for Information Science and Technology*, 51(10), 949-958.
- Rousseau, B.; Rousseau, R. (2000). "Lotka: a program to fit a power law distribution to observed frequency data". *Cybermetrics*, 4(1).
- Rousseau, R. (1997). "Sitations: an exploratory study". *Cybermetrics*, 1(1). <<http://www.cindoc.csic.es/cybermetrics/articles/v1i1p1.html>>. [Consulta: 28/10/2004].
- Rousseau, R. (1999). "Daily time series of common single word searches in AltaVista and NorthernLight". *Cybermetrics*, 2/3. <<http://www.cindoc.csic.es/cybermetrics/articles/v2i1p2.html>>. [Consulta: 28/10/2004].

- Scharnhorst, A. (2004). "Conceptualization of the Web in complex network theory". *4S-EASST Meeting*. Paris, 25-28 agosto.
- Sherman, C.; Price, G. (2001). "The invisible Web. Uncovering information resources search engines can't see". Medford, New Jersey: Information Today Inc.
- Smith, A.; Thelwall, M. (2002). "Web Impact Factors for Australasian universities". *Scientometrics*, 54(3), 363-380.
- Snyder, H.; Rosenbaum, H. (1999). "Can search engines be used as tools for web-link analysis?: a critical view". *Journal of Documentation*, 55(4), 375-384.
- Thelwall, M. (2001a). "The responsiveness of search engine indexes". *Cybermetrics*. 5(1). <<http://www.cindoc.csic.es/cybermetrics/articles/v5i1p1.html>>. [Consulta: 28/10/2004].
- Thelwall, M. (2001b). "A web crawler design for data mining". *Journal of Information Science*, 27(5), 319-325.
- Thelwall, M. (2002a). "A comparison of sources of links for academic Web Impact Factor calculations". *Journal of Documentation*, 58(1), 60-72.
- Thelwall, M. (2002b). "Conceptualizing documentation on the Web: an evaluation of different heuristic-based models for counting links between university web sites". *Journal of the American Society for Information Science and Technology*, 53(12), 995-1005.
- Thelwall, M.; Wilkinson, D. (2004). "Finding similar academic Web sites with links, bibliometric couplings and colinks". *Information Processing and Management*, 40(3), 515-526. [Preprint].
- Van Raan, A. F. J. (2001). "Bibliometrics and Internet: some observations and expectations". *Scientometrics*, 50(1), 59-63.
- Vaughan, L.; Hysen, K. (2002). "Relationship between links to journal web sites and Impact Factors". *Aslib Proceedings*, 54(6), 356-361.
- Vaughan, L.; Thelwall, M. (2004). "Search engine coverage bias: evidence and possible causes". *Information Processing and Management*, 40(4), 693-707.
- Watts, D. J.; Strogatz, S. H. (1998). "Collective dynamics of "small-world" networks". *Nature*, 393, 440-442.
- Wellman, B.; Salaff, J.; Dimitrova, D.; Garton, L.; Gulia, M.; Haythornwaite, C. (1996). "Computer networks as social networks: collaborative work, telework and virtual community". *Annual Review of Sociology*, 22, 213-238.
- WISER. <<http://www.wiserweb.org/>>. [Consulta: 28/10/2004].
- Zakon, R. H. (2003). "Hobbes' Internet Timeline v6.0". <<http://www.zakon.org/robert/internet/timeline/>>. [Consulta: 28/10/2004].
- Zelman, A.; Leydesdorff, L. (2000). "Threaded e-mail messages in self-organization and science and technology studies oriented mailing lists". *Scientometrics*, 48(3), 361-380.