

***AltaVista*, el 'Citation Index' de la Red, y la edad de la inocencia**

***AltaVista*, the 'Citation Index' of the Internet, and the age of innocence**

José-Luis Ortega

Ortega, José-Luis (2022). "*AltaVista*, el 'Citation Index' de la Red, y la edad de la inocencia". *Anuario ThinkEPI*, v. 17, e17e35.

<https://doi.org/10.3145/thinkepi.2023.e17a35>

Publicado en *IweTel* el 19 de octubre de 2023

José-Luis Ortega

<https://orcid.org/0000-0001-9857-1511>

<https://www.directorioexit.info/ficha426>

Instituto de Estudios Sociales Avanzados (IESA-CSIC)
Unidad Asociada Conocimiento, Transferencia e
Innovación (UCO-CSIC), Córdoba
jortega@iesa.csic.es



Resumen: En esta pequeña nota me gustaría rendir homenaje al artículo "Valoración del impacto de la información en Internet: *AltaVista*, el 'Citation Index' de la Red", que Rodríguez-Gairín publicó en 1997 en la *Revista española de documentación científica*. Se trataba de una contribución pionera que ponía de relieve el potencial de los buscadores para generar análisis cuantitativos de enlaces y sitios web. Esta contribución analiza cada una de sus partes, desgranando sus principales aportaciones y discutiendo cuán diferente nos parecía esa Web a la que hoy nos enfrentamos.

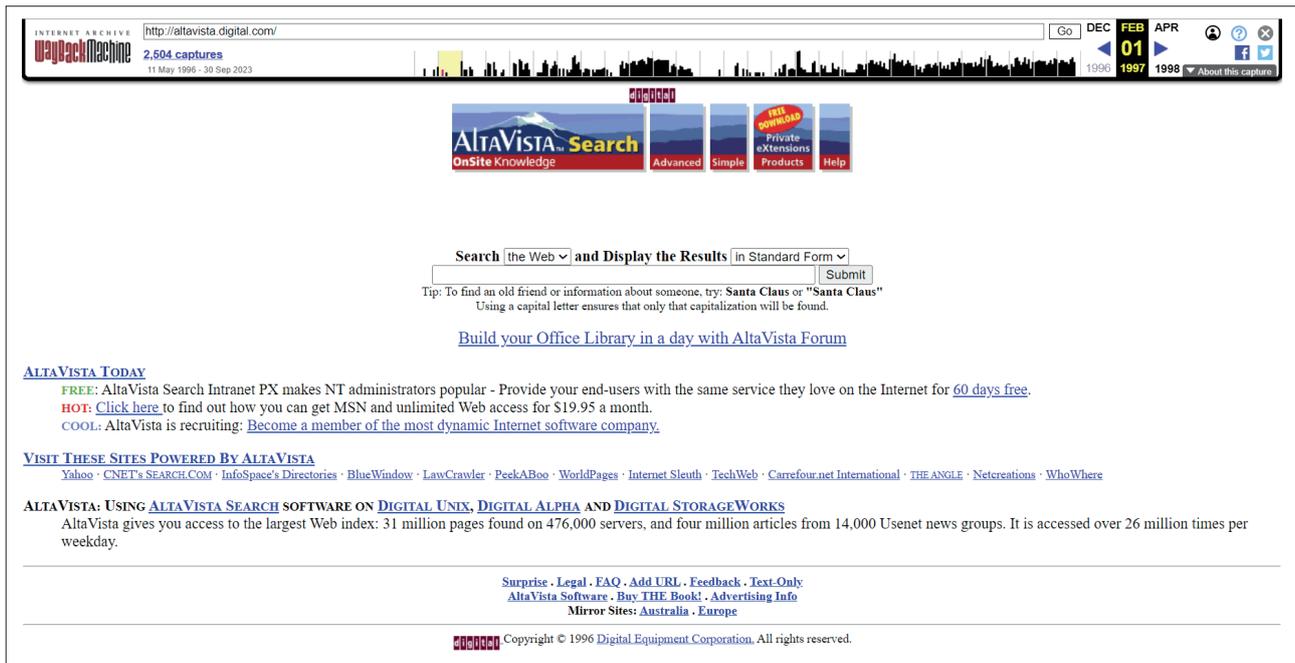
Palabras clave: *AltaVista*; Buscadores; Cibermetría; Meta-etiquetas.

Abstract: This short note attempts to pay tribute to the article "Valoración del impacto de la información en Internet: *AltaVista*, el 'Citation Index' de la Red", published in the *Revista española de documentación científica* in 1997 by Rodríguez-Gairín. The paper was a pioneering contribution regarding the ability of search engines to produce quantitative analyses of links and websites. This work analyses each element and their principal ideas, and discusses how different that Web was from the current version.

Keywords: *AltaVista*; Search engines; Cybermetrics; Metadata.

El artículo

En 1997, la *Revista española de documentación científica* lanzó uno de los artículos más originales e innovadores del momento. Cuando la indización automática, catálogos en línea o la bibliometría descriptiva eran los principales temas en esos años, a Rodríguez-Gairín se le ocurrió señalar el enorme potencial que ofrecían los naciotes buscadores, en concreto *AltaVista*, para cuantificar el impacto y la visibilidad de recursos web. No es que la literatura académica en ese tiempo no hiciese caso a la cada vez más importante gestión y búsqueda de contenido en la Web, sino que hasta ese momento nadie había caído en el potencial de sus operadores y atajos para medir el volumen de contenidos alojados en servidores y dominios, al tiempo que permitían cuantificar el impacto de esos contenidos a través de enlaces entrantes. Y es que "Valoración del impacto de la información en Internet: *AltaVista*, el 'Citation Index' de la Red" (Rodríguez-Gairín, 1997), planteaba por primera vez conectar el mundo de la Web con las crecientes investigaciones en bibliometría, sugiriendo que, si el impacto bibliométrico se basaba en ese momento únicamente en las citas indizadas en el *Science* y *Social Citation Index*, *AltaVista* podría tener el mismo rol para conocer el impacto de los contenidos web a través de sus "sitas" (Rousseau, 1997).



Internet Archive. Home Page de Altavista en febrero de 1997.
<https://web.archive.org/web/19970201010313/http://altavista.digital.com>

El trabajo de Rodríguez-Gairín no era un artículo de investigación, sino una propuesta eminentemente metodológica que exploraba distintos aspectos de la recuperación en la Web. Podemos dividirlo en tres partes bien definidas.

El artículo se inicia comentando cómo la inserción de metadatos en nuestras páginas web ayuda a su indexación en motores de búsqueda. En concreto, las etiquetas meta, hoy consideradas un estándar en el diseño web, empezaban en ese momento a abrirse camino hacia lo que más adelante conoceremos como la web semántica. Con un claro afán pedagógico, el autor nos presenta distintos ejemplos de cómo insertar estas etiquetas con su correspondiente descripción, acompañada de las primeras normalizaciones (*OCLC* y *Dublin Core*).

A continuación, el artículo se sumerge en los recientes operadores y delimitadores que *Altavista* implementó en ese año. Se trataba de una serie de comandos que permitían buscar páginas en función de dónde estaban alojadas y de enlaces de acuerdo con el texto que contenía el URL. Quizás, es aquí donde radica la gran aportación de este trabajo, ya que por primera vez se intuye que usando este tipo de operadores es posible cuantificar el tamaño de la Web visible en un buscador, tanto a nivel de países (*domain:*) como de servidores (*host:*). Además, el delimitador *link:* nos permite filtrar enlaces que apuntan a determinados dominios o sub-dominios, midiendo y contabilizando la visibilidad e impacto de países y sitios web concretos.

Finalmente, Rodríguez-Gairin realiza un ejercicio práctico que nos permite comprobar y entender cómo la combinación de estos operadores permite valorar y comparar la visibilidad e impacto de diversos dominios y sitios web. Así, encontramos una comparación entre dominios de países europeos y luego un gráfico con las principales organizaciones académicas del país. Para ello desarrolla un indicador para normalizar la comparación. El cociente de enlaces externos por el número de páginas, adelantándose así un año al *Web Impact Factor (WIF)* de **Ingwersen** (1998).

Las lecciones

En bibliometría se hace referencia al fenómeno de las “bellas durmientes” para referirse a esos artículos que en un principio pasan desapercibidos, para luego ejercer un fuerte impacto (**Van-Raan**, 2004). El trabajo de Rodríguez-Gairín puede considerarse un buen ejemplo de este tipo, ya que, al estar escrito en castellano y publicado en una revista nacional, resultó desconocido para muchas de las nuevas investigaciones en cibermetría (**Arroyo-Vázquez et al.**, 2005). No fue hasta dos años después, cuando autores como **Smith** (1999a; 1999b) y **Björneborn e Ingwersen** (2001) reconocieran a Rodríguez-Gairín como uno de los primeros precursores de la cibermetría. Gracias también a los seminarios *Cybermetrics* organizados en 1999 por Isidro Aguillo, donde los principales cibérmétricos pudieron tener noticia de este

trabajo. Su primicia residía en que, por sólo unos meses, se adelantase al “Informetric analyses on the World Wide Web: Methodological approaches to ‘Webometrics’” de **Almind e Ingwersen** (1997), trabajo considerado seminal de la cibermetría. La diferencia entre ambos es que mientras Almind e Ingwersen sólo señalaban el potencial para la cibermetría de los nuevos operadores de *AltaVista*, Rodríguez-Gairín hace un ejercicio práctico de estos operadores, sugiriendo lo que a posteriori se conocería como *WIF*.

Y es que este trabajo tan especial nos permite retrotraernos a una época en la que los buscadores todavía mantenían la apariencia de tradicionales bases de datos, con sus operadores booleanos, comandos y búsquedas avanzadas. Una época en la que aún pensábamos que las páginas se podían “catalogar” y los portales temáticos (*Yahoo!*, *DMOZ*) triunfaban al mostrarnos una Web clasificada y disciplinada. En definitiva, pensábamos que la información en la Web podía ser gestionada como gestionábamos el mundo analógico, con catálogos, bibliotecas e índices. En cierto modo, una ilusión en la que todavía no éramos conscientes de que la Web, a la espera de *Google* y la Web 2.0, sería un mundo en sí mismo, regido por sus propias reglas.

Quizás, esta traslación de lo impreso al mundo digital también influyó en la idea que encarnó la cibermetría, una disciplina que nació con fuerza en el crepúsculo del siglo XX y que de forma un tanto inocente creía en una Web como una base de datos, los buscadores como índices de citas y la propia disciplina, una especie de traducción acrítica de la bibliometría a la Web. Sin embargo, esta ingenuidad por desarrollar métricas y metodologías (*crawlers*, *scrapers*) sería fundamental para afrontar los retos actuales de la almetría, redes sociales académicas, la publicación electrónica y la ciencia abierta. Fue una “edad de la inocencia” en la que visionarios como Rodríguez-Gairín se adentraron en una naciente Web inconscientes de cómo esta iba a transformar la ciencia y la sociedad.

Referencias

- Almind, Toas C.; Ingwersen, Peter** (1997). “Informetric analyses on the world wide web: methodological approaches to ‘webometrics’”. *Journal of documentation*, v. 53, n. 4, pp. 404-426.
<https://doi.org/10.1108/EUM0000000007205>
- Arroyo-Vázquez, Natalia; Ortega, José-Luis; Pareja, Víctor; Prieto, José-Antonio; Aguillo, Isidro F.** (2005). “Cibermetría. Estado de la cuestión”. En: *9ªs Jornadas Españolas de Documentación, Fesabid 2005*, Madrid 14-15 de abril.
<https://digital.csic.es/bitstream/10261/4296/1/R-17.pdf>
- Björneborn, Lennart; Ingwersen, Peter** (2001). “Perspective of webometrics”. *Scientometrics*, v. 50, pp. 65-82.
<https://doi.org/10.1023/A:1005642218907>
- Ingwersen, Peter** (1998). “The calculation of web impact factors”. *Journal of documentation*, v. 54, n. 2, pp. 236-243.
<https://doi.org/10.1108/EUM0000000007167>
- Rodríguez-Gairín, Josep-Manuel** (1997). “Valoración del impacto de la información en Internet: *Altavista*, el ‘Citation Index’ de la Red”. *Revista española de documentación científica*, v. 20, n. 2, pp. 175-181.
<https://doi.org/10.3989/redc.1997.v20.i2.591>
- Rousseau, Ronald** (1997). “Sitations: an exploratory study”. *Cybermetrics*, v. 1, n. 1.
<https://dialnet.unirioja.es/descarga/articulo/1256896.pdf>
- Smith, A. G.** (1999a). “ANZAC webometrics: exploring Australasian Web structures”. In: *Proceedings of information online and on disc*, 1999, pp. 159-181.
- Smith, A. G.** (1999b). “The Impact of Web sites: a comparison between Australasia and Latin America”. In: *Proceedings of INFO*, 1999.
- Van-Raan, Anthony F. J.** (2004). “Sleeping beauties in science”. *Scientometrics*, v. 59, n. 3, pp. 467-472.
<https://doi.org/10.1023/B:SCIE.0000018543.82441.f1>