

Web, webometrics, *Wikipedia* e inteligencia artificial

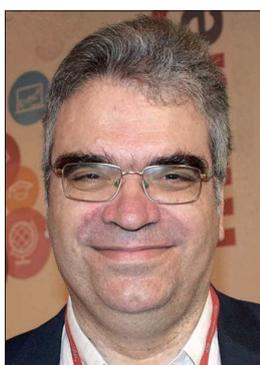
Web, Webometrics, *Wikipedia* and Artificial Intelligence

Isidro F. Aguillo

Aguillo, Isidro F. (2025). "Web, webometrics, *Wikipedia* e inteligencia artificial". *Anuario ThinkEPI*, v. 19, e19a02.

<https://doi.org/10.3145/thinkepi.2025.e19a02>

Publicado en *IweTel* el 11 de febrero de 2025



Isidro F. Aguillo

<https://www.directorioexit.info/ficha67>

<https://orcid.org/0000-0001-8927-4873>

Cybermetrics Lab (IBP-CSIC)

isidro.aguillo@csic.es

Resumen: Posición personal sobre la importancia de generar contenidos abundantes y de calidad en la web pública para mejorar los resultados que proporcionan las herramientas de inteligencia artificial.

Palabras clave: Web; Webometrics; Rankings; *Wikipedia*; Inteligencia artificial.

Abstract: Personal position on the importance of generating abundant and quality content on the public web to improve the results provided by artificial intelligence tools.

Keywords: Web; Webometrics; Rankings; *Wikipedia*; Artificial intelligence.

Al enfrentarse a nuevas situaciones son muy valiosas las experiencias previas que uno ya haya tenido pues pueden proporcionar contexto para enfrentar esas nuevas realidades. Se trata de sugerir consejos o propuestas de buenas prácticas desde la perspectiva que dan los años, especialmente si son ricos en intercambios con maestros y colegas.

Sirva esta introducción para comentar mi experiencia personal con la llegada de internet y la Web y cómo de algunas de las anécdotas de ese periodo podríamos desarrollar estrategias en la era de la inteligencia artificial.

En los primeros años de la Web, en mis cursos de introducción, con frecuencia centrados en los motores de búsqueda (fijación que no he abandonado), siempre tenía que responder a inquietudes de los alumnos/as relativas a la ausencia de contenidos locales en nuestros idiomas. Frente a los que hablaban de colonialismo cultural de los anglosajones o de brecha digital por falta de inversiones o formación, yo defendía una cierta desidia patria en la generación de contenidos autóctonos.

En esos primeros años de la Web era más difícil encontrar información amplia y veraz de "héroes" locales como Viriato o Aníbal que de Hannibal Lecter o de Hannibal Smith (de la serie de TV "El equi-

po A”). Achacaba esa ausencia a una Academia poco interesada en la Web o, como ahora “también” ocurre respecto a la IA, a una inercia de un colectivo decididamente inmovilista.

Algunos recordarán que, con espíritu crítico y motivador, citaba como modelos el portal sobre Gaudí en una Escuela de Arquitectura de Australia (Melbourne creo) antes de que hubiera nada en universidades catalanas, o el episodio de la Guerra del Fletan con Canadá cuando nuestro embajador allí creó el portal web *Si Spain*, una iniciativa de marketing para mostrar aspectos relevantes de nuestro país ante un público poco informado. Aquel portal casi gubernamental fue durante mucho tiempo un referente de información sobre España, labor que como ya sabemos sigue siendo necesaria entre aquellos también norteamericanos que todavía se cuestionan en que parte de México nos encontramos y si somos morenos de raza árabe.

Pero mi mejor anécdota deriva de otra serie de TV de aquellos años: En “Las chicas de oro” aparecía el personaje de Rose (interpretado por la magnífica Betty White), que con frecuencia mencionaba su lugar de origen en San Olaf en Minnesota. La serie la presentaba como una persona simple de una ciudad que el folklore local parece identificar como un lugar no muy avanzado. Pues bien, en aquel entonces, el St. Olaf College, la universidad local, tenía una sección de páginas web en español que era un inesperado referente en nuestro idioma.

No solo era preocupación por la ausencia de contenidos lo que motivaba estos comentarios en mis cursos. Ya entonces era evidente que no todas las fuentes tenían la misma calidad y que, de hecho, en algunos temas sensibles como los relacionados con la salud, los contenidos eran incorrectos, peligrosos o falsos. En esos momentos no había iniciativas contra las *fake news*, por lo que adopté una aproximación diferente. Describía las sedes web académicas como “islas de calidad” en un archipiélago que también incluía páginas de la llamada internet invisible (catálogos de biblioteca y bases de datos bibliográficas y otras bases curadas de otro tipo). Por tanto, había que encontrar un medio de enriquecer los contenidos de esas fuentes tanto en volumen como en calidad.

Encontré que un arma poderosa para tal fin era realizar rankings basados en indicadores que midieran precisamente tanto la cantidad de contenidos web como su calidad. En un escenario competitivo, la lucha por posiciones de cabecera podría animar a la difusión en abierto de los conocimientos generados por las instituciones académicas y de investigación. Los rankings web (Webometrics) de universidades, centros de investigación, hospitales y repositorios han incrementado de manera sustancial los contenidos web de calidad, entre otros, en muchos países del llamado Sur Global.

La irrupción de las redes sociales supuso un cierto varapalo al fin de hacer públicos contenidos de calidad. No es solo cierto que en la mayoría de los casos la información que proporcionan es superficial o volátil (la adquisición de *Twitter/X* por Elon Musk muestra que puede ser sesgada, interesada o directamente falsa), sino que también afecta a los contenidos de las sedes web institucionales o personales. Muchas sedes universitarias utilizan CMS poco flexibles, priorizando la rigidez burocrática y el control, sobre la libertad de publicación de contenidos dinámicos lo que afecta en particular a las páginas personales. Sedes externas como blogs o proyectos concretos han dejado paso a mensajes cortos o *tiktoks* y en muchos casos han terminado desapareciendo.

¿Qué tiene que ver esto con la inteligencia artificial? Pues mucho, porque son los contenidos públicos los que sirven para entrenar los modelos LLM y por tanto de la disponibilidad de los mismos y de su calidad va a depender la idoneidad y relevancia de las respuestas que proporcione la IA. Es cierto que también se pueden utilizar otras fuentes, pero estas suelen ser más controladas y por tanto la arena está en la web pública. La actual generación de chatbots no siempre ofrece información exacta, se dice que alucina, cuando posiblemente trata simplemente de llenar lagunas de conocimiento.

Lagunas que generan problemas serios que me retrotraen a los casos que he presentado anteriormente. Generar contenidos en abierto será clave.

Hace poco se ofreció un *prompt* para generar una entrada personal en *Wikipedia*. Vanidoso, utilicé mi propio caso para probar en alguno de los servicios de IA descubriendo un resultado bastante amplio y fidedigno (aparte de halagador). Sin embargo, analizando en más detalle la respuesta descubrí que la mayor parte de los datos provenían de información disponible en inglés y que textos clave en español habían sido ignorados. Es de esperar que esto mejore en el futuro y que nuestros textos encuentren su camino en la IA.

Y ya que he mencionado *Wikipedia*, mi última reflexión tiene que ver con la evolución de la percepción que se ha tenido de esta herramienta. Al igual que con la IA, la *Wikipedia* fue recibida al principio con bastante escepticismo, cuando no rechazo. Se cuestionaba la calidad de los contenidos y en especial cierta opacidad de la autoría. Sin embargo, con el tiempo se ha reconocido no solo el valor informativo de los contenidos, sino especialmente la cita de fuentes primarias en cada entrada. Ese es el valor añadido que proporciona la *Wikipedia* y que en sistemas de IA más sofisticados, como por ejemplo *Perplexity*, ya se ofrece por defecto.

Como colofón, sigo dando prioridad a la publicación en abierto de contenidos académicos, especialmente los generados por el sector público. Y más allá de infraestructuras rígidas, promover la participación individual con libertad de creación y volver a una web más rica y diversa. Esa es la fuente con la que debemos potenciar en la dirección correcta lo que nos pueda aportar la inteligencia artificial.

Nota: Este trabajo ha sido redactado por IA (Isidro Aguillo) sin contribución alguna de IA (inteligencia Artificial).