LOS PERIÓDICOS ESPAÑOLES COMO FUENTE DE REFERENCIA EN *WIKIPEDIA*

Spanish newspapers as a reference source of *Wikipedia*

David Rodríguez-Mateos y Tony Hernández-Pérez

✉ **David Rodríguez-Mateos**, doctor en Documentación, es profesor del *Departamento de Periodismo y Comunicación Audiovisual* de la *Universidad Carlos III de Madrid* y miembro del grupo *TecnoDoc*. Imparte docencia e investiga sobre tecnología audiovisual, gestión de contenidos web, *e-learning*, *web scraping* y documentación periodística y audiovisual.
https://orcid.org/0000-0003-1555-5685

pirio@bib.uc3m.es

Tony Hernández-Pérez es doctor en Ciencias de la Información y profesor del *Departamento de Biblioteconomía y Documentación* de la *Universidad Carlos III de Madrid*. Su labor docente e investigadora está ligada al grupo *TecnoDoc* incluyendo asignaturas de web social, gestión de contenidos web, metadatos, búsqueda y recuperación de información, ciencia abierta, y documentación periodística y audiovisual.
https://orcid.org/0000-0001-8404-9247

tony@bib.uc3m.es

Universidad Carlos III de Madrid
Facultad de Humanidades, Comunicación y Documentación
Madrid, 128. 28903 Getafe (Madrid), España

Resumen

En España *Wikipedia* ocupa el noveno lugar entre los sitios webs más consultados. Las consultas más frecuentes son sobre: personajes (deportistas, artistas, políticos, etc.); películas y series de televisión; lugares; o deportes, entre otros. Muchos de estos temas requieren una actualización casi continua. La política oficial de *Wikipedia* en español declara que "las fuentes especializadas deben tener prioridad sobre la prensa generalista, ya que diarios y semanarios tienden a simplificar su exposición, omitiendo a veces datos importantes o describiendo el tema de forma enciclopédicamente poco útil". El artículo analiza hasta qué punto *Wikipedia* respeta esta política analizando la presencia en ella de algunos de los medios de comunicación españoles más importantes como fuente de referencia externa. ¿Cuantos artículos, y qué porcentaje, de la Wikipedia en español poseen referencias o citas a los periódicos digitales españoles más consultados en internet? ¿Sobre qué tipos de contenidos de *Wikipedia*, desde el punto de vista temático, se emplean los periódicos digitales españoles como fuente externa?

Palabras clave

Periódicos; Prensa digital; Diarios digitales; *Wikipedia*; Citas; Referencia; Fiabilidad; Fuentes de información.

Abstract

Spanish *Wikipedia* ranks ninth among the most consulted websites. The most frequent queries are about: celebrities (athletes, artists, politicians, etc.); movies and television series; places; or sports, among others. Many of these issues require an almost continuous update. The official policy of the *Wikipedia* declares that "specialized sources should have priority over the general press, since newspapers and weekly magazines tend to simplify their exposition, sometimes omitting important data or describing the subject in an encyclopedically unhelpful way". The article analyzes the extent to which *Wikipedia* respects this policy by analyzing the presence of some of the most important Spanish newspaper media in *Wikipedia* as a source of external reference. How many articles, and what percentage, of the Spanish *Wikipedia* have references or citations to the Spanish digital newspapers most consulted on the internet? On what types of contents of *Wikipedia*, from the thematic point of view, are Spanish digital newspapers used as an external source?

Keywords

Newspapers; Digital press; Digital newspapers; *Wikipedia*; Cites; Reference; Information sources; Reliability.

Artículo recibido el 05-04-2018
Aceptación definitiva: 03-10-2018

Rodríguez-Mateos, David; **Hernández-Pérez, Tony** (2018). "Los periódicos españoles como fuente de referencia en *Wikipedia*". *El profesional de la información*, v. 27, n. 6, pp. 1323-1333.

https://doi.org/10.3145/epi.2018.nov.15

1. Introducción

Wikipedia ocupa el quinto lugar entre los sitios webs más visitados del mundo y el noveno en España, según datos de *Alexa*.
https://www.alexa.com/topsites/countries/ES

Su creador Jimmy Wales, junto con otros miembros de la comunidad, elaboró entre 2001 y 2002 una serie de reglas con el fin de definir, estructurar, legitimar y aumentar la credibilidad de la información que se difundía.

Estas reglas son conocidas como los "cinco pilares" (*Wikipedia*, 2018):

- *Wikipedia* es una enciclopedia, por lo que los contenidos no son originales, sino que están basados en otros;
- busca un punto de vista neutral, lo que significa no presentar ningún punto de vista como el verdadero o el mejor, y eso supone

  > "citar fuentes autorizadas que puedan verificarse siempre que sea posible";

- es de contenido libre, por lo que cualquiera puede tanto acceder al contenido como editarlo;
- sigue unas normas de etiqueta: no todos los posibles editores tienen la misma capacidad para corregir textos, y las correcciones deben seguir unas ciertas pautas;
- no tiene otras normas firmes, aparte de las anteriores.

Los cinco pilares han conseguido que *Wikipedia* siga siendo muy consultada y utilizada, incluso en la Academia, como lo demuestran las estadísticas de tráfico web y diversos estudios. Por ejemplo, **Aibar** *et al.*, (2015) señalan que estudiantes y profesores de universidad valoran positivamente la calidad de los resultados y la usan regularmente, o **Azzam** *et al.*, (2017) muestran que profesores y estudiantes universitarios usan *Wikipedia* para mejorar la calidad de la información sobre medicina que se difunde a través de esta enciclopedia.

2. Medios de comunicación y *Wikipedia*

La política oficial de *Wikipedia* en español declara que

> "las fuentes especializadas deben tener prioridad sobre la prensa generalista, ya que diarios y semanarios tienden a simplificar su exposición, omitiendo a veces datos importantes o describiendo el tema de forma enciclopédicamente poco útil, al dirigirse a un público más amplio que no tiene por qué dominar la disciplina" (*Wikipedia*, 2017).

Y que

> "las noticias procedentes de las agencias de prensa o medios de contrastada reputación sí que permiten referenciar correctamente asuntos de actualidad cuando no haya otra fuente para tal efecto. A este respecto, es conveniente no olvidar que la misma noticia en distintos medios puede ser tratada de forma completamente diferente, incluso contradictoria" (*Wikipedia*, 2017),

lo cual debe hacerse constar claramente para respetar el punto de vista neutral sobre el tema.

Sin embargo, el uso más frecuente de muchos de los usuarios habituales de *Wikipedia* no es precisamente el de la comunicación científica, sino que fundamentalmente se consulta sobre:

- personajes: deportistas, músicos, artistas, políticos, nobles y personas físicas o jurídicas famosas;
- películas y series de televisión;
- lugares: países, regiones, localidades;
- períodos históricos: desde días concretos como Navidad o Santos Inocentes a II Guerra Mundial;
- deportes;
- conceptos: tabla periódica de elementos, *bitcoin*, halo, contaminación, globalización, etc.;
- medicina: síndrome de asperger, naproxeno, etc.,
  https://tools.wmflabs.org/topviews/?project=es.wikipedia.org&platform=all-access&date=last-month&excludes=

Una parte de los artículos relativos a este tipo de información contiene referencias no de fuentes de información científicas sino de los medios de comunicación, tanto de prensa generalista como de prensa especializada. Así pues, los artículos de *Wikipedia* en español citan como fuente autorizada y verificable las informaciones publicadas en los medios de comunicación tradicionales.

3. En busca de la calidad: la importancia de las referencias en *Wikipedia* como indicio de fiabilidad

La fiabilidad de *Wikipedia*, en términos de precisión y tamaño, sigue estando entre los temas más recurrentemente investigados, con algunos detractores destacados (**Sahut**; **Tricot**, 2017; **Hube**, 2017; **Viseur**, 2014; **Giles**, 2005). Una de las principales críticas estudiadas por **Tramullas**, **Garrido-Picazo** y **Sánchez-Casabón** (2016), es la facilidad con la que cualquiera puede editar y manipular información, el vandalismo en la *Wikipedia*.

A diferencia de las enciclopedias tradicionales, en donde la redacción y edición de los artículos estaba en manos de reputados profesores de la Academia, en la *Wikipedia* la responsabilidad de la redacción y edición de los artículos recae en el público en general, lo que ha sido señalado por algunos como un signo del triunfo de la mediocridad y la popularidad sobre el criterio de la ciencia y de los expertos:

> "En la internet, sin embargo, se encomia e incluso venera el amateurismo en lugar de la experiencia. Hoy, el *Oxford English Dictionary* y la *Encyclopaedia Britannica*, dos obras de referencia en las que hemos confiado durante

> mucho tiempo para obtener información, están siendo reemplazadas por la *Wikipedia* y otros recursos generados por los usuarios. Se reemplaza al profesional por el aficionado, al lexicógrafo por el laico, al profesor de Harvard por la población no escolarizada" (**Keen**, 2007).

No es la única crítica. También hay estudios sobre el papel de los editores, que muchas veces se convierten en protagonistas de las llamadas "guerras de ediciones", causadas a veces por un sesgo ideológico o simplemente por una imposición de criterios sobre contribuidores menos experimentados (**Jirschitzka** *et al.*, 2017; **Baeza-Yates**; **Sáez-Trumper**, 2015; **Yasseri** *et al.*, 2012), que alejan a nuevos posibles editores, lo que a su vez puede provocar el problema de una enciclopedia con cada vez menos contribuyentes.

A pesar de las críticas, el público la percibe como una buena fuente de información, de ahí su elevada consulta. Es evidente que existe un gran interés sobre cómo evaluar la calidad de los artículos en *Wikipedia*. **Shen**, **Qi** y **Baldwin** (2017) hacen un buen resumen de la bibliografía existente. Estos trabajos suelen ser básicamente de tres tipos:

- estudio de los metadatos de los artículos: autoridad de los editores, redes de editores, número de ediciones, evolución de las distintas ediciones, etc.;
- estudio de la estructura de los artículos: longitud de un artículo, ratio de referencias respecto a la longitud, número de encabezamientos, número de palabras, estilo de escritura, etc;
- estudios híbridos: analizan la relación entre los metadatos y la estructura interna de los artículos.

Wikipedia dispone de un sistema de clasificación que permite determinar la calidad de los artículos de la enciclopedia (*Wikipedia*, 2015) y en español distingue una escala que va desde los artículos destacados a artículos que son apenas un esbozo. La escala determina si un artículo tiene calidad:

- destacada;
- buena;
- media;
- aceptable;
- baja;
- esbozo.

Para que un artículo sea destacado debe cumplir ciertos criterios (*Wikipedia*, 2016):

- basado en fuentes fiables;
- aseveraciones verificables mediante referencias;
- presente un punto de vista neutral;
- bien escrito y presentado;
- completo, extenso y profundo;
- cumpla con el manual de estilo y la estructura de un artículo;
- sea estable, que no sufra guerra de ediciones ni su contenido cambie frecuentemente.

Por todo ello el uso de referencias a fuentes externas en los artículos se ha generalizado, ya que

> "las citas [a referencias externas] proporcionan credibilidad y fomentan la confianza, enlazan contenido a fuentes de conocimiento e instituciones existentes (…) proporcionan una autoridad externa" (**Sahut**; **Tricot**, 2017).

Y ello a pesar de que durante los primeros años de vida de *Wikipedia* no se citaban fuentes de referencia. De hecho, la institucionalización de la práctica de referenciar fuentes en los artículos generó entre 2004 y 2008 mucho debate entre la comunidad de wikipedistas en Francia (**Sahut**, 2014) y también en España, en donde se pueden observar algunas discusiones de 2009 en la página en español de *Wikipedia* sobre fuentes fiables (*Wikipedia*, 2017).

El uso de fuentes externas ha llegado a motivar estudios, no sólo sobre la calidad de las referencias sino incluso sobre el impacto de las referencias científicas existentes en ella. Así, **Kousha** y **Thelwall** (2017) intentan comprobar si *Wikipedia* podría proporcionar nuevas evidencias sobre el impacto de la investigación académica, para lo que toman las referencias a 302.328 artículos y 18.735 monografías en inglés indexados por *Scopus* en el período de 2005 a 2012, y **Serrano-López**, **Ingwersen** y **Sanz-Casado** (2017) analizan las referencias a artículos académicos sobre energía eólica en artículos de *Wikipedia*.

Lo que nos interesa en este artículo, entre otras cosas, es averiguar si los artículos de *Wikipedia* citan como referencia, como fuente fiable, a algunos de los principales periódicos españoles. Y si lo hacen, en qué medida, qué periódicos y para qué temas. El análisis de los datos podrá revelar si los editores de la enciclopedia consideran los periódicos una fuente fiable.

> "En *Wikipedia* las citas en los artículos proporcionan credibilidad, fomentan la confianza, enlazan contenido a fuentes de conocimiento e instituciones existentes y proporcionan una autoridad externa"

4. Objetivos y metodología

El artículo trata de analizar la presencia en *Wikipedia* de algunos de los medios de comunicación españoles más importantes como fuente de referencia externa con el fin de posteriormente estimar su posible influencia en la *Wikipedia* española. Los objetivos y las preguntas de investigación que se pretenden responder son:

- ¿Cuantos artículos, y qué porcentaje, de *Wikipedia* en español poseen referencias a los periódicos digitales españoles más consultados en internet?
- ¿Sobre qué tipos de contenidos de *Wikipedia*, desde el punto de vista temático, se emplean los periódicos digitales españoles como fuente externa?
- ¿Tienen presencia, se citan como fuente de referencia los medios de comunicación analizados en los artículos "destacados" y "buenos" de *Wikipedia*?

La investigación inicial se llevó a cabo sobre el contenido de todos los artículos incluidos en la *Wikipedia* en español. *Wikipedia* compila cada cierto período de tiempo todo su contenido en diversos formatos, incluyendo tanto los artículos en sí como la suma de éstos con las plantillas, los objetos multimedia, etc.

En este caso, se ha optado por la versión compilada más

reciente al inicio de la investigación. Esta compilación se completó el 1 de diciembre de 2017, es de 11.426 MB, 206.093.404 líneas, una vez descargado en formato XML para su procesamiento y está accesible a través de la web: *https://archive.org/details/eswiki-20170201*

El fichero contiene cuatro tipos de informaciones:

- 1.370.602 artículos. Cada artículo es una entrada de *Wikipedia*. Ejemplo: "Andorra", *https://es.wikipedia.org/wiki/Andorra*
- plantillas empleadas para realizar artículos;
- descripciones de ficheros de imágenes y audiovisuales;
- meta-páginas primarias.

Para esta investigación únicamente se han tenido en cuenta los artículos. La extracción y tabulación inicial de los datos se realizó mediante *Python*, basándose en el siguiente flujo de trabajo:

1) Extraer el contenido de cada artículo.

2) Extraer su url.

3) Obtener dentro del artículo sus categorías.

4) Extraer del texto del artículo las urls relativas a alguno de los medios analizados.

5) Elaborar con toda la información obtenida una tabla de resultados con:

- nombre del artículo;
- todas las referencias a medios periodísticos, incluyendo el nombre del medio y la url empleada;
- todas las categorías de *Wikipedia* empleadas para clasificar el artículo;
- todas las categorías en las que se incluye la entrada.

> *Wikipedia* dispone de un sistema de clasificación para determinar la calidad de los artículos, un tema que preocupa mucho a la comunidad

En este trabajo se considera que una url es relativa a un medio concreto si coinciden al menos en los dominios de primer nivel y de segundo nivel. Ello se debe a que un mismo medio puede tener más de un dominio de tercer nivel. Un ejemplo: dos urls que empiecen por "www.elpais.es" y "retina.elpais.es" se considerarán como pertenecientes a un mismo medio. La primera de ellas pertenecerá probablemente a los contenidos más generalistas del medio, mientras que la segunda pertenece a uno de sus suplementos específicos, que tendrá incluso una distinción tipográfica a primera vista, pero que forma parte de los contenidos.

Tabla 1. Lista de periódicos digitales con mayor audiencia en España en agosto, septiembre y octubre de 2017 según datos de ComScore y OJD Interactiva (OK diario, 2017)

Medio	Septiembre 2017		Agosto 2017		Julio 2017	
	Ranking (visitas)	Visitas (000)	Ranking (visitas)	Visitas (000)	Ranking (visitas)	Visitas (000)
Elmundo.es	1	169.536	2	161.747	1	148.276
Elpais.com	2	162.942	1	162.243	2	144.994
Lavanguardia.com	3	81.258	3	78.264	3	62.265
Abc.es	4	65.838	4	66.717	5	60.733
Elconfidencial.com	5	63.735	5	59.569	4	61.437
20minutos.es	6	36.646	6	39.705	6	35.407
Okdiario.com	7	34.749	7	32.331	7	30.335
Elperiodico.com	8	28.114	8	27.493	11	20.115
Eldiario.es	n/d	n/d	9	25.755	8	24.414
Elespanol.com	10	23.204	11	22.410	10	22.803
Lavozdegalicia.es	11	21.891	10	23.902	9	23.875
Huffingtonpost.es	n/d	n/d	12	21.126	12	19.607
Libertaddigital.com	13	18.898	14	15.088	14	14.029
Publico.es	14	18.730	13	18.213	13	16.311
Ideal.es	15	n/d	15	12.316	16	11.687

Cada enlace de los artículos fue revisado para cotejar si pertenecía a la versión digital de alguno de los 18 periódicos de mayor audiencia en España. La pertenencia a este grupo estaba basada en los datos medidos por *ComScore* y *OJD Interactiva*, de forma consecutiva, durante los meses de julio, agosto y septiembre de 2017, como se muestra en la tabla 1. Originalmente se pensó tomar como referencia los 20 periódicos más leídos, aunque se observó que alguno de ellos aparecía en varios de los meses, pero no en todos. Se decidió elegir a aquellos que al menos aparecieran en dos de esos tres meses.

Finalmente se obtuvieron todos los artículos marcados como destacados y buenos por *Wikipedia* en español, de acuerdo con los datos ofrecidos por la propia *Wikipedia* (véase un ejemplo en la figura 1). Para cada uno de ellos se obtuvo el número total de referencias mediante un nuevo script de *Python*.

No obstante, estas referencias no están siempre en la misma sección de *Wikipedia* (p. ej., "Referencias"), y no siempre están marcadas internamente con el mismo código, por lo que se realizó una revisión manual, al menos en aquellos registros con un bajo número de referencias obtenidas automáticamente.

5. Referencias a periódicos en *Wikipedia*

La versión de *Wikipedia* con la que se trabajó, publicada con fecha 1 de diciembre de 2017, contenía 1.370.602 artículos y 4.661.800 enlaces en total. De estos artículos, 39.810 (2,9%) contenían referencias a alguno de los 18 medios analizados:

- algunas veces, una o varias referencias a un único medio de los 18;

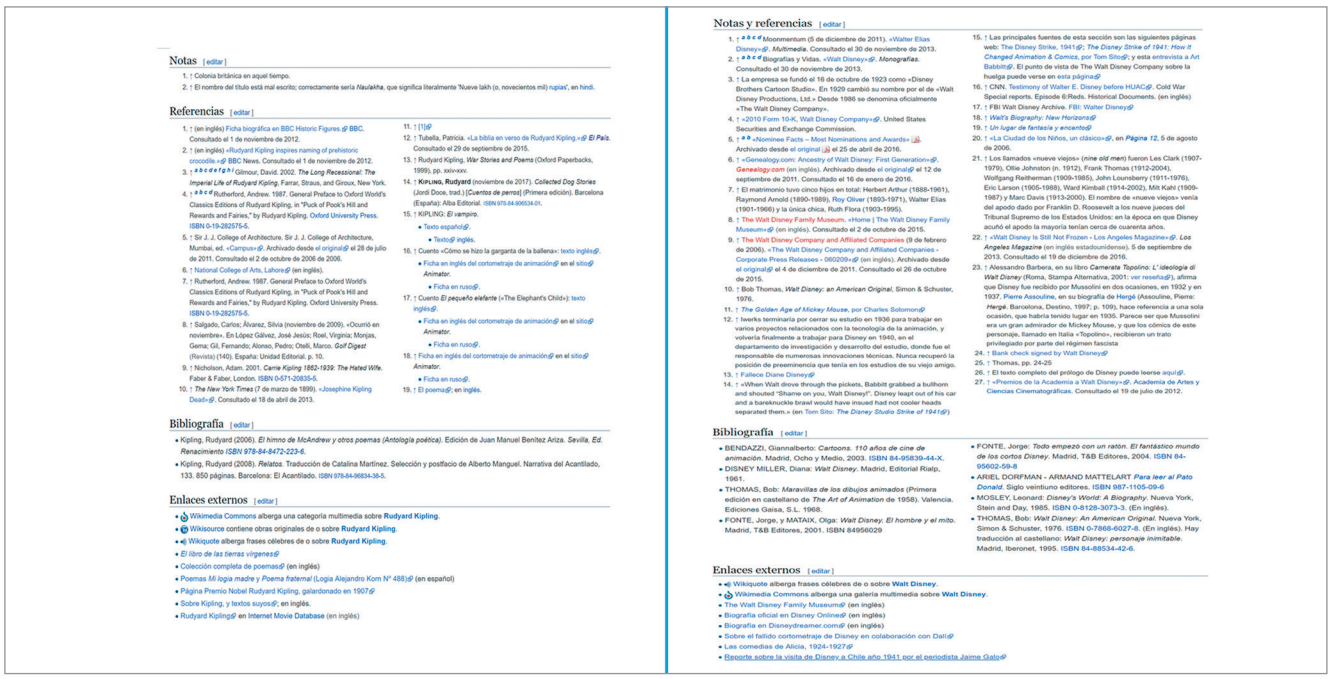


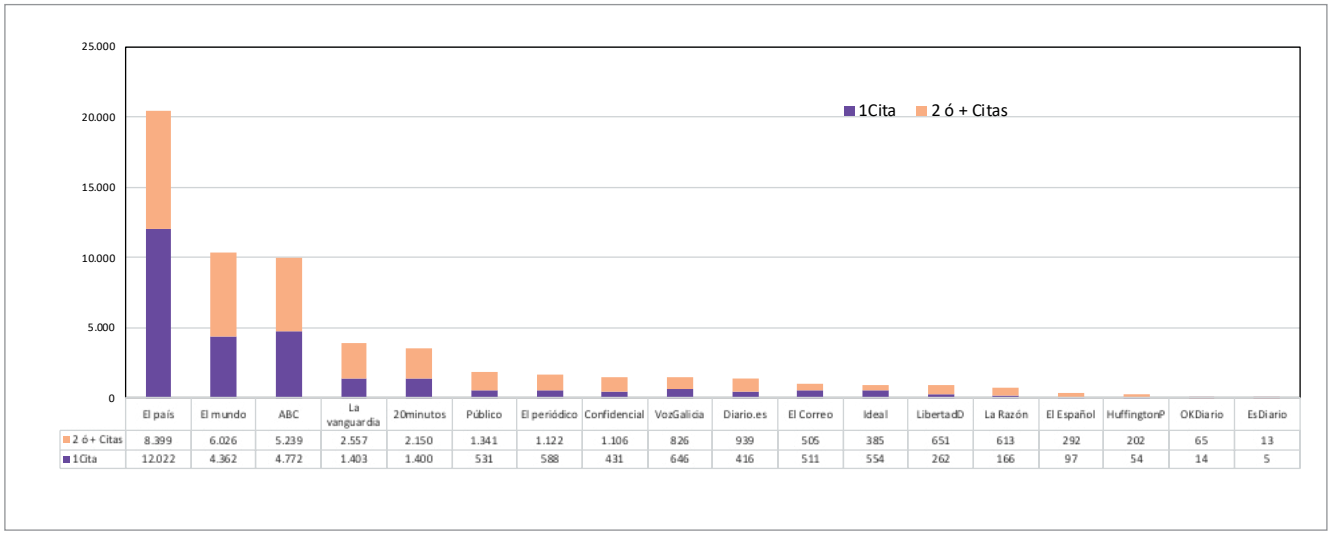
Figura 1. Dos ejemplos de las referencias externas citadas en los artículos de Wikipedia sobre Rudyard Kipling (izquierda) y Walt Disney (derecha). Fecha de acceso: 28/3/2018.
https://es.wikipedia.org/wiki/Rudyard_Kipling
https://es.wikipedia.org/wiki/Walt_Disney

- otras, al menos dos referencias a dos medios, se supone que con el fin de contribuir a buscar el punto de vista neutral y a aumentar la calidad y la percepción de credibilidad en el lector.

Si analizamos esos 39.810 artículos que contienen referencias a los 18 medios, podemos observar que los cinco primeros: *El país*, *El mundo*, *ABC*, *La vanguardia* y *20 minutos*, aparecen en el 79,66% de todos los artículos con referencias a alguno de los 18 medios de comunicación estudiados. El análisis del número de enlaces (a medios o no) incluidos en estos artículos arroja datos parecidos: estos cinco medios acaparan el 82,45% de los 92.519 enlaces que contienen los artículos.

La figura 2 muestra los artículos que incluyen referencias a alguno de los medios incluidos. El total es mayor que el citado número de artículos (39.810), ya que un mismo artículo puede incluir referencias a más de un medio.

No obstante, en contra de lo que se podría esperar, la mayoría de los artículos (28.235, un 70,9%) contiene referencias a un solo medio de comunicación de los 18, mientras que 11.575 (29,1%) contienen referencias, al menos, a dos medios (en el caso más extremo, un artículo tiene referencias hasta a 15 medios). En ambos casos, artículos con referencias a un solo medio o a más de uno, *El país* es el medio más citado y su presencia en *Wikipedia* duplica al segundo, *El mundo*. Es más, el número de referencias a *El país* es prácti-



	El país	El mundo	ABC	La vanguardia	20minutos	Público	El periódico	Confidencial	VozGalicia	Diario.es	El Correo	Ideal	LibertadD	La Razón	El Español	HuffingtonP	OKDiario	EsDiario
2 ó + Citas	8.399	6.026	5.239	2.557	2.150	1.341	1.122	1.106	826	939	505	385	651	613	292	202	65	13
1Cita	12.022	4.362	4.772	1.403	1.400	531	588	431	646	416	511	554	262	166	97	54	14	5

Figura 2. Artículos de *Wikipedia* en español que contienen alguna cita (una, o más de una) a alguno de los periódicos incluidos en el estudio

Tabla 2. Artículos de *Wikipedia* con referencia a un mayor número de medios

Nº de medios de la muestra con referencias en el artículo	Artículo
15	Unión Progreso y Democracia
14	Atentados de Cataluña de 2017
14	Abdicación de Juan Carlos I
14	Lionel Messi
14	Podemos
14	Referéndum de independencia de Cataluña de 2017
13	Elecciones al Parlamento de Andalucía de 2015
13	Proceso soberanista de Cataluña de 2012-2017
13	Anexo: Fallecidos en abril de 2017
13	Anexo: España en 2013
13	Corrupción urbanística en España
13	Anexo: España en 2015

camente el mismo que la suma de las referencias a *El mundo* y al diario *ABC*.

El caso de *El país* es especialmente llamativo en el número de artículos que contienen una sola cita a los medios citados, 12.022 artículos, que equivale a la suma de los cinco siguientes medios. Esta afirmación no significa que los artículos tengan una sola referencia y que sea a *El país*, sino que puede haber referencias a otros sitios web o incluso a periódicos extranjeros. Lo que nos dice este dato es que en este caso sería el único medio español de los 18 estudiados que cuenta con una referencia en esos artículos concretos.

La tabla 2 muestra los títulos de los 12 artículos con mayor diversidad de referencias de fuentes de medios de comunicación. No significa necesariamente que sean los artículos sobre los que más se ha disputado o los más polémicos sino aquellos en los que hay una mayor presencia de medios analizados en sus textos.

Como era de esperar son temas de marcado carácter político con la excepción de la figura de un deportista: Lionel Messi. El análisis de los 100 primeros artículos por número de medios citados arroja resultados muy parecidos: una preponderancia absoluta (más del 80%) de artículos sobre política (partidos, situación de España, crisis política y económica), seguido por una proporción mucho menor de personas (fallecidos por meses, deportistas, actores o personajes de la televisión) o de algún asunto internacional (refundación de la *UE*, guerra contra el Estado Islámico o la *Jornada Mundial de la Juventud 2011*).

El análisis sobre cómo estos artículos comparten referencias de los medios (figura 3), muestra que *El país*, *El mundo* y *ABC* son los diarios que más referencias comparten. Además de con *El mundo* y *ABC*, *El país* comparte también un relativamente alto número de referencias con *Público* y *La vanguardia*; y en menor medida con *El periódico* y *20 minutos*. Por su parte *El mundo* y *ABC* tienden a compartir más referencias, además de con *El país*, con *El confidencial* y *La vanguardia*, siendo el resto de los casos más marginales.

6. Temas relacionados con las referencias: categorías

Una de las grandes ventajas de *Wikipedia* es que sus artículos pueden ser asignados a categorías, lo cual es una de forma de agrupar artículos; también se pueden agrupar por listas o por plantillas de navegación. El objetivo principal de las categorías es agrupar páginas para proporcionar enlaces de navegación de forma jerárquica, de tal forma que un usuario pueda descubrir otras páginas sobre un tema a partir de alguna de las características que conozca de ese tema.

Por ejemplo, el "aceite de oliva" es una categoría en sí misma en *Wikipedia*, de la que dependen tres subcategorías:

- aceite de oliva con denominación de origen;
- aceituna;
- alfarería de aceite.

Además permite descubrir otros términos relacionados con esas categorías y subcategorías (almazara, *Consejo Oleícola Internacional*, olivicultura, etc.). Al asignar categorías a un artículo, *Wikipedia* genera de forma automática un enlace a ese artículo en su categoría, lo que permite buscar mediante navegación (*browsing*) en vez de mediante la caja de búsqueda.

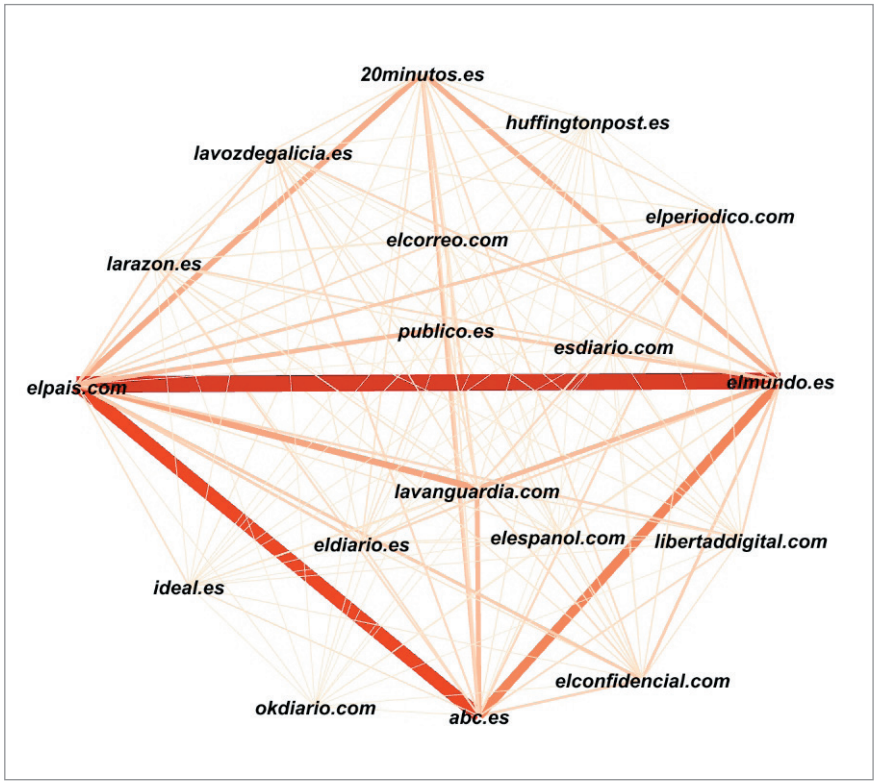


Figura 3. Relación entre referencias a distintos medios en mismos artículos de *Wikipedia*

Tabla 3. Categorías de *Wikipedia* con mayor número de artículos que incluyen referencias a medios

Categorías de *Wikipedia*	Nº de artículos con referencias a medios
Escritores de España del siglo XX	729
Escritores de España del siglo XXI	631
Escritores en español del siglo XX	625
Actores de televisión de España	596
Madrileños	586
Actores de cine de España	516
Escritores en español del siglo XXI	495
Actores de teatro de España	467
Presentadores de televisión de España	420
Políticos de España del siglo XX	401
Escritores en español	315
Medalla de Oro al Mérito en las Bellas Artes	283
Políticos de España del siglo XXI	277
Montevideanos	267
Premio *Ondas*	263
Periodistas de España del siglo XX	258
Caballeros grandes cruces de la *Orden de Isabel la Católica*	255
Políticos del *Partido Socialista Obrero Español*	250
Políticos del *Partido Popular*	248
Pintores de España del siglo XX	247

No todo son ventajas. *Wikipedia* está llena de ejemplos de problemas de "sobrecategorización":

- asignar a un artículo muchas categorías, una de las cuales contiene a la otra;
- creación innecesaria de categorías con poco sentido;
- asignación de múltiples categorías a un artículo.

Por ejemplo, el artículo sobre Federico García Lorca tiene 28 categorías. La propia *Wikipedia* tiene varias páginas dedicadas al tema. Y como es sabido, la definición y el mantenimiento de las categorías es difícil.
https://es.wikipedia.org/wiki/Wikipedia:Categor%C3%ADas,_listas_y_plantillas_de_navegaci%C3%B3n
https://bit.ly/2z5VSGi

No obstante, las categorías ofrecen la posibilidad de analizar en qué tipos de artículos aparecen más referencias de los medios de comunicación y por tanto conocer los campos en los que los medios están siendo más influyentes. Los 39.810 artículos de la muestra aparecían asociados a 63.274 categorías propias de *Wikipedia*, de las cuales el 98% estaban asociadas a menos de 10 artículos. Más aún, 8 de cada 10 categorías son mencionadas únicamente en un pequeño número de artículos (entre 1 y 3).

> En *Wikipedia* las citas o referencias a fuentes externas no están bien normalizadas: no aparecen en el mismo sitio, no se marcan con el mismo código

Un primer análisis de los datos permite deducir que con gran diferencia, existe un mayor número de referencias a periódicos españoles en los artículos de *Wikipedia* relativos a personas y personajes relacionados con las artes, el entretenimiento y el deporte: escritores, actores, actrices, directores de cine, productores, guionistas, presentadores de programas, periodistas, etc. La tabla 3 muestra como ejemplo las 20 categorías que contienen un mayor número de artículos con referencias a los medios analizados.

Para hacer una comparación más en detalle, se ha tomado una muestra de 138 categorías (aquellas que tienen 75 o más artículos con referencias a medios), y se han realizado algunos cálculos estadísticos mediante el software *R* (*R*

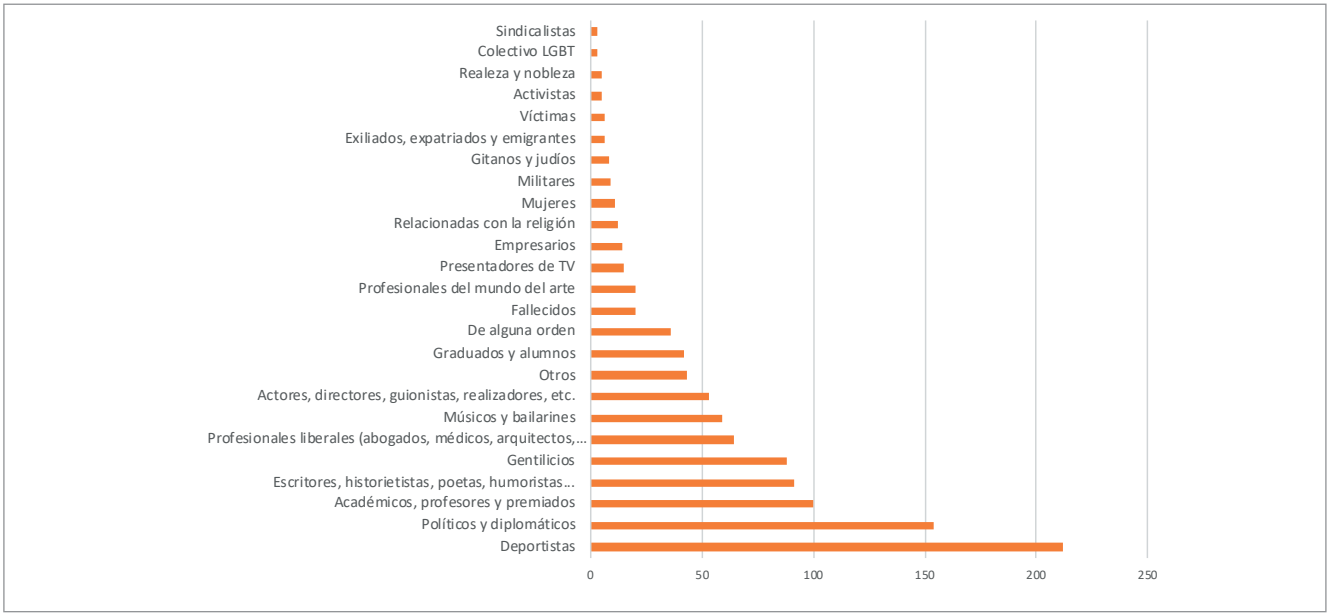


Figura 4. Macrocategorías de personajes

Core Team, 2018). De estas categorías, 65 están relacionadas con personajes del arte, el entretenimiento, el deporte o el periodismo: llamaremos a este grupo de categorías "artes"; las otras 73 categorías corresponden a temas diferentes, y llamaremos a este segundo grupo "no artes".

Para cada uno de estos grupos, se aplicó un test de *Shapiro-Wilk* (**Shapiro**; **Wilk**, 1965): por un lado, cuántos artículos incluían referencias a medios ("con referencias"), y asimismo, cuantos artículos no las incluían ("sin referencias").

De sus resultados se deduce que ninguna de las dos distribuciones de categorías es normal, como resume la tabla 4.

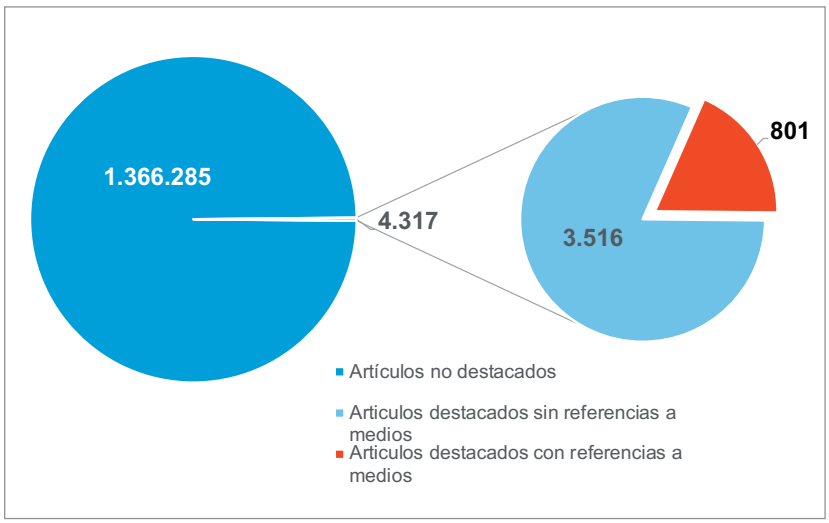


Figura 5. Artículos destacados o buenos y presencia de los medios analizados

Se ha realizado un test de rangos o test de Wilcoxon (**Wilcoxon**, 1945) a la muestra de artículos de la categoría "artes", comparando en este caso los artículos con y sin referencias (V=67,5, p<0,05). De ello se podría inducir que hay diferencias significativas en ambas muestras, en favor de las categorías relacionadas con "artes".

Un análisis adicional se ha realizado extrayendo las 1.080 primeras categorías de *Wikipedia* relacionadas con personas, como se muestra en la figura 4. Se han agrupado en conjuntos más genéricos de macrocategorías, con el fin de obtener tendencias sobre tipos de contenidos. De este análisis se puede deducir que son los artículos sobre deportistas (de distintas especialidades) los que con mayor frecuencia contienen un gran número de referencias a periódicos españoles.

Gentilicios se refiere a personas pertenecientes a algún lugar (madrileños, catalanes).

De alguna orden son personas que pertenecen a una orden militar, religiosa, etc. (por ejemplo, *Orden de Malta*, *Orden de Santiago*, etc.).

Profesionales del mundo del arte son pintores, escultores, etc.

La categoría Otros es necesariamente amplia dada la diversidad de categorías existentes (por ejemplo, zurdos, vegetarianos, aviadores, personas centenarias, etc.).

Gitanos y judíos se refiere a categorías como gitanos españoles, judíos sefardíes, judíos de Francia, de Argentina, etc.

En cuanto a los artículos destacados o buenos, *Wikipedia* cuenta con 4.317 artículos (un 0,31%) del total, lo cual muestra que la consideración de "destacado" o "bueno", según los propios editores de *Wikipedia* en español, es muy reducida. Sin embargo si analizamos de esos artículos destacados o buenos aquellos en los que aparecen referencias a alguno de los medios analizados, aparecen 801 artículos destacados o buenos, es decir, un 18,55% del total, según muestra la figura 5.

Tabla 4. Test de Shapiro aplicado a una muestra de artículos con y sin referencias a medios que están incluidos en categorías "artes", o bien, que no están incluidas en esas categorías

	Artes	No artes
Con referencias a medios	W = 0,67999, p<0,05	W = 0,57765, p<0,05
Sin referencias a medios	W = 0,64001, p<0,05	W = 0,70837, p<0,05

Finalmente, también se ha analizado el total de referencias (tanto a medios como a otras fuentes) de todos los artículos destacados y buenos, como se muestra en la figura 6. En este caso se detecta que la presencia de referencias a medios coincide con un mayor número de referencias, tanto considerando los artículos destacados por separado como los buenos y la suma de ambos, como señala la figura.

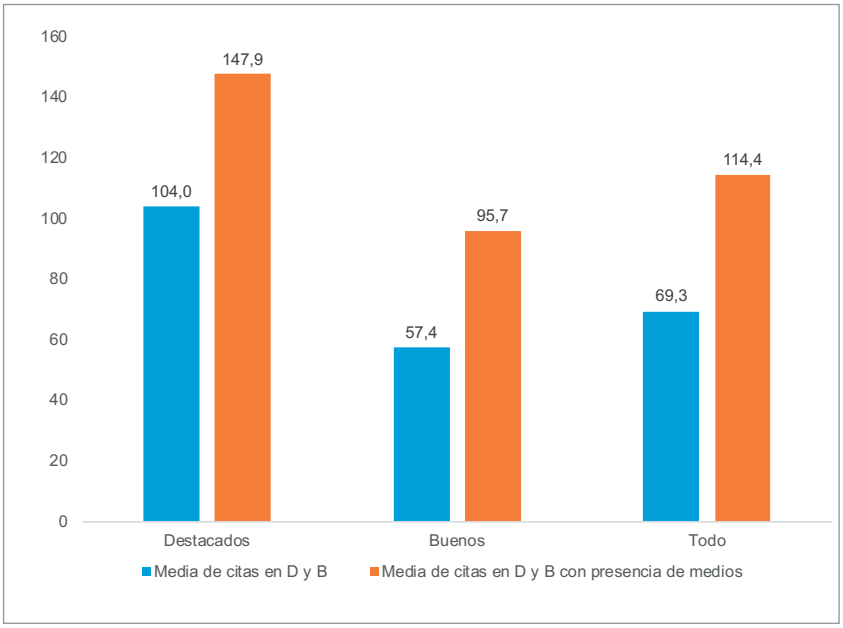


Figura 6. Media de referencias en artículos destacados y buenos (D y B) de *Wikipedia*, y relación con la presencia de medios

Adicionalmente se ha realizado una prueba de ji-cuadrado con corrección de continuidad de Yates (**Yates**, 1984), partiendo de la hipótesis (H_0) de que no hubiera diferencias significativas de artículos buenos y destacados en los artículos con referencias a medios, o sin referencias a ellos. Para ello se realizó inicialmente una tabla de contingencia con las referencias obtenidas en los artículos destacados y buenos, como se indica en la tabla 5.

A partir de los datos de esa tabla, los resultados (ji-cuadrado = 3.432,9, df=1, p<0,05) señalarían que sí existe una asociación significativa entre ciertos artículos, que son considerados como buenos o destacados, y la existencia en los mismos de referencias a medios periodísticos.

7. Discusión

Los medios de comunicación españoles más importantes se encuentran representados en la bibliografía o en las referencias de los artículos de *Wikipedia* únicamente en un 2,9% (39.810 artículos), lo cual resulta coherente con los propios principios de la enciclopedia, que considera que las fuentes especializadas deben tener prioridad sobre la prensa generalista.

Con diferencia *El país* es el medio de comunicación más citado como referencia en la *Wikipedia* española: está citado en 20.421 artículos y mediante 35.557 enlaces, seguido de:

- *El mundo*, en 10.338 artículos
- *ABC*, 10.011
- *La vanguardia*, 3.960
- *20 minutos*, 3.550

Estos cinco diarios suman prácticamente el 80% de todas las referencias a los medios de comunicación españoles analizados. Si se observa el gráfico 2 se observa que en 28.234 ocasiones (70,92% de los artículos), la referencia es a un solo medio, lo que puede contradecir con el principio de búsqueda del punto de vista neutral de *Wikipedia*.

> Apenas un 3% de los artículos de *Wikipedia* en español contiene citas a los diarios españoles más importantes

Debe tenerse en cuenta que cuando decimos "referencia a un solo medio" queremos decir a uno de los estudiados. Por ejemplo, en el artículo de *Wikipedia* sobre Alan García, la única referencia a un medio de comunicación español es *El país*, pero existen referencias a *El mercurio* y a otros diarios peruanos. En la página sobre "energía solar", la única referencia a un medio español es de nuevo a *El país*, pero existen también referencias a la *Agencia Reuters*.

La mayoría de las referencias pertenecen a medios que disponen tanto de versión impresa como digital (los cinco citados). En el caso de *El confidencial*, a pesar de estar entre los cinco diarios más leídos (en digital), su presencia en *Wikipedia* es meramente testimonial.

En cuanto a la temática, las referencias en *Wikipedia* a contenidos publicados en los medios de comunicación son más frecuentes de lo esperado en cierto tipo de artículos: sin duda, las personas físicas –deportistas, escritores, o personas y personajes del mundo del espectáculo en general- y jurídicas –empresas y organizaciones de todo tipo- son, con diferencia, los que más referencias a medios de comunicación hacen. Tiene su lógica, resulta más difícil encontrar referencias de fuentes especializadas de todo lo que tiene que ver con la actualidad y, por tanto, muchas veces se acude a reportajes o entrevistas a estos personajes.

Tabla 5. Tabla de contingencia con las referencias a artículos de *Wikipedia* considerados como destacados o buenos, que contengan o no referencia a los medios analizados

Referencias en artículos	Tipos de artículos		Total de referencias
	Destacados	Buenos	Total
Sin medios	72.703	134.946	207.649
Con medios	42.456	49.207	91.663
Total general	115.159	184.153	299.312

> *Wikipedia* prefiere que se referencien las fuentes especializadas sobre la prensa generalista

Del mismo modo, los artículos sobre deportes en general requieren de una constante actualización cuya fuente son muchas veces los medios de comunicación. También los términos geográficos (países y secundariamente ciudades) tienen mucha presencia en las referencias a los medios de comunicación que hacen los artículos de *Wikipedia*. La explicación es muy parecida: la enciclopedia tiende a estar muy actualizada en cuanto a datos económicos, geográficos, políticos…, y los medios suelen cubrir muy bien este tipo de información. No hemos estudiado si con el tiempo y la aparición de fuentes de referencia más especializadas se tiende a modificar estas referencias y bibliografía.

La temática de los artículos cambia sustancialmente cuando hay referencias a más de un medio. La tabla 2 muestra cómo cuantas más referencias a medios hay en un artículo, más probabilidades existe de que el artículo sea de política: la visión de España, independencia de Cataluña, partidos políticos, etc. Con la excepción de Messi y algunos listados de personas (fallecidos en algún año, por ejemplo), la gran mayoría de los artículos en donde hay al menos dos referencias a medios de comunicación tienen que ver con asuntos políticos y económicos.

Merece la pena destacar que casi el 20% de los artículos considerados de buena calidad incluye la presencia de alguna cita a los medios analizados, muy por encima de lo que cabría esperar. Asimismo, el volumen medio de referencias de los artículos con mayor calidad es también mayor por término medio cuando esas referencias incluyen medios.

8. Conclusiones

A principios de diciembre de 2017, *Wikipedia* en español contenía 1.370.602 artículos y 4.661.800 enlaces en total. De estos artículos, 39.810, apenas un 2,9% del total contenían referencias a alguno de los medios de comunicación

analizados. No se puede considerar una cifra elevada y significa que la versión española de *Wikipedia* no toma a los medios de comunicación españoles como una fuente de referencia principal, tal y como establece la propia política oficial de la enciclopedia.

Personas y personajes de todo tipo representan la temática más importante en los artículos de *Wikipedia*, especialmente cuando el artículo contiene referencias a un único medio. Cuando un artículo comparte referencias a múltiples medios de comunicación la temática más habitual suele ser la política.

> En *Wikipedia* en español *El país* es con diferencia el diario con más citas

Si bien las referencias a los medios representan apenas el 2,9% de los artículos de *Wikipedia*, la proporción sube hasta el 18,55% en los artículos destacados y buenos, lo que supone un reconocimiento importante al valor de estas referencias en la enciclopedia.

Con diferencia, *El país* es el medio más representado en las referencias de los artículos de *Wikipedia*. Y junto con las referencias de *El mundo*, *ABC*, *La vanguardia* y *20 minutos* representan el 80% de todas las referencias en artículos. No creemos que eso denote un sesgo ideológico-político evidente hacia las ideas de ese medio puesto que cuando se tratan temas políticos y económicos rara vez aparece citado como único medio, de lo que también se deduce una preocupación real por el respeto al punto de vista neutral del que hace gala *Wikipedia*.

Por último, la presencia de las referencias a los medios es mayor de lo esperado en los artículos con mayor calidad, si consideramos al menos dos factores para definirla:

- la evaluación subjetiva de los propios editores de *Wikipedia* (que deciden si un artículo es "destacado" o "bueno");
- el número medio de referencias de esos artículos.

9. Referencias

Aibar, Eduard; **Lladós-Masllorens, Josep**; **Meseguer-Artola, Antoni**; **Minguillón, Julià**; **Lerga, Maura** (2015). "Wikipedia at university: What faculty think and do about it". *The electronic library*, v. 33, n. 4, pp. 668-683.
https://bit.ly/2EPstq0
https://doi.org/10.1108/EL-12-2013-0217

Azzam, Amin; **Bresler, David**; **Leon, Armando**; **Maggio, Lauren**; **Whitaker, Evans**; **Heilman, James**; **Orlowitz, Jake**; **Swisher, Valerie**; **Rasberry, Lane**; **Otoide, Kingsley**; **Trotter, Fred**; **Ross, Will**; **McCue, Jack D.** (2017). "Why medical schools should embrace Wikipedia: Final-year medical student contributions to Wikipedia articles for academic credit at one school". *Academic medicine*, v. 92, n. 2, p. 194.
https://doi.org/10.1097/ACM.0000000000001381

Baeza-Yates, Ricardo; **Sáez-Trumper, Diego** (2015). "Wisdom of the crowd or wisdom of a few? An analysis of users' content generation". In: *Procs of the 26th ACM Conf on hypertext & social media*, pp. 69-74.
https://doi.org/10.1145/2700171.2791056

Giles, Jim (2005). "Internet encyclopaedias go head to head". *Nature*, n. 438, pp. 900-901.
https://doi.org/10.1038/438900a

Hube, Christoph (2017). "Bias in Wikipedia". In: *Procs of the 26th Intl Conf on world wide web companion*, pp. 717-721.
https://doi.org/10.1145/3041021.3053375

Jirschitzka, Jens; **Kimmerle, Joachim**; **Halatchliyski, Iassen**; **Hancke, Julia**; **Meurers, Detmar**; **Cress, Ulrike** (2017). "A productive clash of perspectives? The interplay between articles' and authors' perspectives and their impact on Wikipedia edits in a controversial domain". *PLoS one*, v. 12, n. 6, e0178985.
https://doi.org/10.1371/journal.pone.0178985

Keen, Andrew (2007). *The cult of the amateur: how today's internet is killing our culture*. New York: Doubleday. ISBN: 978 0 385 52080 5
https://filmadapter.files.wordpress.com/2014/10/andrew_keen_the_cult_of_the_amateur_how_todaysbookfi-org.pdf

Kousha, Kayvan; **Thelwall, Mike** (2017). "Are Wikipedia citations important evidence of the impact of scholarly articles and books?". *Journal of the Association for Information Science and Technology*, v. 68, n. 3, pp. 762-779.
https://doi.org/10.1002/asi.23694

OK diario (2017). "comScore. Nuevo récord de visitas de *Okdiario*: 39.380.408 y ya es sexto en el Top 10 de periódicos generalistas". *Okdiario*, 20 octubre.
https://okdiario.com/audiencia/2017/10/20/comscore-septiembre-2017-1433604

R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
https://www.R-project.org

Sahut, Gilles (2014). "'Citez vos sources': archéologie d'une règle au cœur du savoir wikipédien (2002-2008)". *Études de communication. Langages, information, médiations*, n. 42, pp. 97-110.
https://doi.org/10.4000/edc.5721

Sahut, Gilles; **Tricot, André** (2017). "Wikipedia: An opportunity to rethink the links between sources' credibility, trust, and authority". *First Monday*, v. 22, n. 11.
https://ojphi.org/ojs/index.php/fm/article/view/7108

Serrano-López, Antonio-Eleazar; **Ingwersen, Peter**; **Sanz-Casado, Elías** (2017). "Wind power research in Wikipedia: Does Wikipedia demonstrate direct influence of research publications and can it be used as adequate source in research evaluation?". *Scientometrics*, v. 112, n. 3, pp. 1471-1488.
https://doi.org/10.1007/s11192-017-2447-2

Shapiro, Samuel-Sanford; **Wilk, Martin-Bradbury** (1965). "An analysis of variance test for normality (Complete samples)". En: *Biometrika*, v. 52, n. 3-4, pp. 591-611.
http://www.bios.unc.edu/~mhudgens/bios/662/2008fall/Backup/wilkshapiro1965.pdf

Shen, Aili; **Qi, Jianzhong**; **Baldwin, Timothy** (2017). "A hybrid model for quality assessment of Wikipedia articles". In:

Procs of the Australasian Language Technology Association workshop 2017, pp. 43-52.
http://aclweb.org/anthology/U17-1005

Shenoy, Aravind; **Prabhu, Anirudh** (2016). "Ranking in SEO". In: Shenoy, Aravind; Prabhu, Anirudh. *Introducing SEO*, pp. 21-35. ISBN: 978 1 4842 1854 9
https://doi.org/10.1007/978-1-4842-1854-9_3

Tramullas, Jesús; **Garrido-Picazo, Piedad**; **Sánchez-Casabón, Ana I.** (2016). "Research on Wikipedia vandalism: a brief literature review". In: *Procs of the 4th Spanish conference on information retrieval*, p. 15.
https://arxiv.org/pdf/1606.05609

Viseur, Robert (2014). "Reliability of user-generated data: The case of biographical data in Wikipedia". In: *Procs of The intl symposium on open collaboration*, p. 31.
https://bit.ly/2JkprIN
https://doi.org/10.1145/2641580.2641618

Wikipedia (2015). "Plantilla:Sistema de clasificación". *Wikipedia, la enciclopedia libre*.
https://es.wikipedia.org/w/index.php?title=Plantilla:Sistema_de_clasificaci%C3%B3n&oldid=80225389

Wikipedia (2016). "Wikipedia:Qué es un artículo destacado". *Wikipedia, la enciclopedia libre*.
https://bit.ly/2Jpc76g

Wikipedia (2017). "Wikipedia:Fuentes fiables". *Wikipedia, la enciclopedia libre*.
https://es.wikipedia.org/w/index.php?title=Wikipedia:Fuentes_fiables&oldid=97853047

Wikipedia (2018). "Wikipedia:Los cinco pilares". *Wikipedia, la enciclopedia libre*.
https://es.wikipedia.org/wiki/Wikipedia:Los_cinco_pilares

Wilcoxon, Frank (1945). "Individual comparisons by ranking methods". *Biometrics bulletin*, v. 1, n. 6, pp. 80-83.
https://sci2s.ugr.es/keel/pdf/algorithm/articulo/wilcoxon1945.pdf

Yasseri, Taha; **Sumi, Robert**; **Rung, András**; **Kornai, András**; **Kertész, János** (2012). "Dynamics of conflicts in Wikipedia". *PLoS one*, v. 7, n. 6, e38869.
https://doi.org/10.1371/journal.pone.0038869

Yates, Frank (1984). "Tests of significance for 2 × 2 contingency tables". *Journal of the Royal Statistical Society*, v. 147, n. 3, pp. 426-463.
http://mathfaculty.fullerton.edu/sbehseta/Yates-twobytwo.pdf

20 años diseñando y gestionando información

MASmedios apoya la Declaracion de Lyon del 2014 que propugna el derecho de las personas a acceder a la información.

masmedios

www.masmedios.com