

Análisis del contenido científico de la Wikipedia española

2016

Eduard Aibar, Julià Minguillón, Maura Lerga, Josep Lladós, Antoni Meseguer, Peter Dunajcsik

Open Science & Innovation Research Group (<https://osi.blogs.uoc.edu/>)



Análisis del contenido científico de la Wikipedia española por E. Aibar, J. Minguillón, M. Lerga, J. Lladós, A.

Meseguer y P. Dunajcsik se distribuye bajo una [Licencia Creative Commons Atribución-CompartirIgual 4.0 Internacional](https://creativecommons.org/licenses/by-sa/4.0/).

Índice

1.	INTRODUCCIÓN.....	1
	<i>ALGUNOS DATOS BÁSICOS SOBRE WIKIPEDIA</i>	1
2.	EL CONTENIDO CIENTÍFICO Y TECNOLÓGICO DE LA WIKIPEDIA ESPAÑOLA.....	3
3.	ESTUDIO DE LA CALIDAD DEL CONTENIDO CIENTÍFICO DE LA WIKIPEDIA ESPAÑOLA	6
3.1.	INTRODUCCIÓN	6
3.2.	ESTUDIO CUALITATIVO DE REFERENCIAS EN WIKIPEDIA.....	9
3.3.	PRESENCIA DE REFERENCIAS DE REVISTAS ACADÉMICAS	12
3.4.	ESTUDIO CUANTITATIVO DE REFERENCIAS EN WIKIPEDIA.....	17
4.	ESTUDIO DE LA PRESENCIA DE LA CIENCIA ESPAÑOLA EN WIKIPEDIA.....	21
4.1.	ESTUDIANTES MATRICULADOS.....	21
4.2.	PRODUCCIÓN CIENTÍFICA ESPAÑOLA.....	23
4.3.	PRESENCIA DE INSTITUCIONES CIENTÍFICAS E INVESTIGADORES	29
5.	ESTUDIO DE LOS EDITORES Y DEL PROCESO DE EDICIÓN	41
5.1.	ESTUDIO DE LOS EDITORES DE ARTÍCULOS CIENTÍFICOS DE LA WIKIPEDIA ESPAÑOLA	41
5.2.	ESTUDIO DEL PROCESO DE EDICIÓN	51
6.	CIENCIA Y CONTEXTO SOCIAL.....	54
6.1.	ESTUDIO DE LAS PÁGINAS DE DISCUSIÓN	55
6.2.	CONSTRUCCIÓN DE MAPAS COGNITIVOS	59
7.	CONCLUSIONES.....	63
8.	REFERENCIAS	66

1. Introducción

El objetivo general de este estudio es llevar a cabo un análisis exhaustivo del contenido científico de la versión española de Wikipedia. Por “contenido científico” entendemos todos aquellos artículos de la enciclopedia sobre materias, conceptos, técnicas, etc. perteneciente claramente a alguna disciplina científica – en las distintas áreas de conocimiento, desde las ciencias naturales a las ciencias sociales o las ingenierías.

Como tal el proyecto de investigación constituye el primer estudio sistemático hecho nunca sobre el contenido científico de cualquier versión lingüística de Wikipedia – incluyendo la inglesa que es sobre la que se han generado más estudios y análisis.

De forma más precisa, el proyecto se ha desarrollado alrededor de cuatro ejes de estudio: (1) un estudio de la calidad del contenido científico de la Wikipedia española; (2) un estudio de la presencia de la ciencia española en la Wikipedia en español; (3) un estudio de los editores y del proceso de edición de los artículos en temas de ciencia, de la Wikipedia española; y (4) una aproximación al contexto social de la ciencia en un entorno colaborativo como es la Wikipedia.

Este informe expone, en primer lugar, el proceso llevado a cabo para determinar el conjunto de páginas de ciencia y tecnología de la Wikipedia española, denominado como Corpus de ciencia y tecnología, que ha servido de base para la mayoría de los análisis llevados a cabo en el proyecto.

Algunos datos básicos sobre Wikipedia

Wikipedia, "la enciclopedia libre que todos pueden editar", fue creada en el año 2001. La versión inglesa de Wikipedia tiene más de 5 millones de artículos y la española más de 1.2: unas cifras que la sitúan muy por encima de cualquier otro proyecto enciclopédico anterior. Existen, de hecho, más de 280 versiones lingüísticas de Wikipedia y para muchas de estas lenguas Wikipedia se ha convertido en la primera enciclopedia que han tenido nunca.

Actualmente es la séptima página web más visitada en Internet - sólo superada por Google, Facebook, YouTube, Yahoo, Baidu y Amazon. La versión inglesa recibe más de 500 millones de visitantes únicos cada mes y tiene un promedio de unos 18.000 millones de páginas vistas. Además, y a diferencia de estos otros grandes portales de Internet, Wikipedia no es propiedad de ninguna empresa privada ni de ninguna gran corporación multinacional. Hay que ir hasta la posición 75 del ranking de Alexa de webs más visitadas, para encontrar otra que tampoco lo sea: la web de la BBC, en este caso, una empresa pública del gobierno británico.

Wikipedia ha generado una inmensa cantidad de literatura de todo tipo, desde ensayos académicos, obras divulgativas, numerosas noticias en la prensa e, incluso, una gran cantidad de artículos científicos de base empírica que analizan diversos aspectos de su funcionamiento (Mesgari et al., 2014). Una parte significativa de esta literatura ha dedicado grandes esfuerzos a teorizar sobre Wikipedia y a tratar de entender el sentido y alcance de un fenómeno tan peculiar. Wikipedia ha sido entendida, por poner algunos ejemplos, como instancia

paradigmática de la producción entre iguales basada en el procomún (Benkler, 2006), como modelo por antonomasia de las llamadas aplicaciones web 2.0 – un concepto fuertemente discutido por su ambigüedad –, o como culminación del movimiento open source (Weber, 2004).

Un aspecto de Wikipedia que, en cambio, tiende a pasar mucho menos desapercibido es su uso como fuente de información científica. Diferentes estudios recientes sobre comunicación y percepción pública de la ciencia coinciden en constatar que Internet se ha convertido, actualmente, en la principal fuente de información científica para la mayoría de ciudadanos (Brossard & Scheufele, 2013; FECYT, 2012). En los últimos años Internet ha sobrepasado en este terreno a los medios de comunicación tradicionales: prensa escrita, radio y televisión. Según un estudio de la National Science Foundation (EEUU) más del 60% de los ciudadanos que buscan información científica sobre algún tema específico, recurren en primer lugar a Internet, mientras que sólo un 12% recurren a la versión en línea de medios tradicionales - prensa diaria o revistas (National Science Board, 2012).

Los datos de la Encuesta de Percepción Social de la Ciencia en España de la FECYT (2012) muestran que Internet es la principal fuente de información científica para el público español - en concreto para el 40,9% de los encuestados. Por primera vez en España, Internet se sitúa por encima de la TV (31%) y muy por encima del resto. Preguntados por el tipo de recursos en Internet que utilizan para obtener información científica, el 21,7% manifiesta utilizar Wikipedia como fuente prioritaria. Sólo blogs y redes sociales están por encima, pero dado que tanto unos como otros engloban una gran diversidad de instancias, Wikipedia se convierte de hecho en la fuente singular más consultada y, por tanto, en el canal de comunicación pública de la ciencia más importante en la actualidad (Aibar, 2015).

A pesar de este volumen tan considerable de información científica y de la situación de preponderancia incontestable de Wikipedia como canal de comunicación pública de la ciencia, hay poca conciencia de este hecho - y de sus consecuencias - por parte de los actores sociales más directamente implicados. Por un lado las instituciones de I + D, en particular las universidades y centros de investigación, centran sus esfuerzos de comunicación pública en otros canales al margen de Wikipedia (en el terreno de Internet, por ejemplo, priorizan el uso de redes sociales privadas) y, lo que es aún más preocupante, los mismos científicos y profesionales de la ciencia - incluyendo el profesorado universitario - viven mayoritariamente de espaldas a esta realidad y apenas se implican en la edición o mejora de los contenidos científicos de Wikipedia (Aibar et al., 2015). Muy pocos de ellos son de hecho conscientes del papel preponderante de Wikipedia como canal de comunicación pública de la ciencia y, por tanto, de un elemento crucial en la configuración y mejora de la cultura científica de la sociedad.

2. El contenido científico y tecnológico de la Wikipedia española

Es muy difícil determinar, ni siquiera en términos cuantitativos, cuál es el contenido "científico" y "tecnológico" de Wikipedia, es decir, el subconjunto de artículos de una versión lingüística determinada que tratan temas de ciencia y tecnología. Al problema previo e independiente de determinar lo que se considera ciencia o tecnología, hay que añadir la dificultad aún mayor generada por las características propias de la categorización de los artículos en Wikipedia.

Todos los artículos de la Wikipedia están clasificados mediante categorías que se pueden encontrar en la parte inferior del texto del artículo. Las categorías son asignadas según los criterios de los editores y, como cualquier otro contenido del artículo, están sujetos a la revisión y modificación posterior por parte de otros editores. En la Wikipedia no se cumple, sin embargo, ninguno de los dos principios básicos de las clasificaciones tradicionales: la *exclusividad* y la *exhaustividad* (Salah et al., 2012). La exclusividad es el hecho de que ningún ítem pertenezca a dos categorías diferentes; en Wikipedia un mismo artículo pertenece a una media de 3 a 5 categorías diferentes. La exhaustividad se refiere a que una categoría englobe todos y sólo los ítems que son apropiados; en Wikipedia las categorías no agrupan todos los artículos realmente relevantes y no es extraño que incluyan, además, otros que no lo son.

Las categorías de Wikipedia, por último, no responden a ninguna estructura jerarquizada dado que no siguen el objetivo previo de establecer un sistema rígido de particiones entre grupos de artículos, ni de definir relaciones claras entre las clases. Se trata, más bien, de una clasificación basada en folksonomías (Voss, 2006). Todo ello hace que un artículo como 'Oxígeno' en la versión española, esté categorizado como 'Oxígeno', 'anfígenos', 'bioelementos', o como 'Medicamento Esencial de la Organización Mundial de la Salud', entre otras categorías. Resulta, por tanto, muy problemático utilizar las categorías para delimitar el subconjunto de artículos de ciencia y tecnología.

Dado que las categorías no sirven como taxonomía precisa y que no es posible etiquetar a mano un millón de páginas, en nuestro proyecto de investigación hemos optado por un procedimiento, inspirado en el trabajo de Groves et al. (2015), que crea grupos de páginas fuertemente enlazadas entre sí para determinar las categorías de unos pocos centenares de componentes, los cuales contienen decenas de miles de páginas cada uno. La base de datos original proviene de la descarga de un *dump*¹ de Wikipedia en diciembre de 2014. Mediante la instalación de Python 3 y las Media Utilities² se fueron generando y depurando distintos ficheros de datos conteniendo los enlaces internos entre páginas de la Wikipedia española.

La fase de depuración de datos supuso múltiples tareas intermedias como eliminar redirecciones, ordenar y eliminar duplicados, descartar enlaces a imágenes o categorías, etc. Se tomaron algunas decisiones para delimitar el conjunto final de datos, como descartar las páginas que no tuvieran un tamaño mínimo de bytes u otros conjuntos de contenidos particulares como las páginas de usuario, los anexos o los portales.

¹ <http://dumps.wikimedia.org/eswiki/latest/eswiki-latest-pages-articles.xml.bz2>

² <https://pythonhosted.org/mediawiki-utilities/index.html>

Finalmente, utilizamos el algoritmo *Infomap*³ para generar un fichero compuesto por comunidades y subcomunidades. *Infomap* crea comunidades temáticas a partir del análisis de los enlaces internos entre todos los artículos de Wikipedia contenidos en la base de datos. Cada comunidad es, pues, un conjunto de páginas fuertemente enlazadas entre sí. En un primer lugar, obtuvimos un total de 1.251 comunidades, con más de quinientas mil páginas de Wikipedia en su versión española.

La primera depuración de estos datos supuso la eliminación de enlaces rotos, enlaces inexistentes, páginas sin entrada, corregir los errores derivados de la escritura (problemas con mayúsculas y minúsculas, nombres compuestos escritos con guiones bajos, etc.). Este proceso nos redujo el conjunto inicial a 974 comunidades. La segunda depuración de datos consistió en eliminar aquellos conjuntos de páginas que aparecían repetidos bajo distintos nombres (p.e. páginas de referencia), obteniendo una lista definitiva de 458 comunidades, que agrupaban un total de 205.907 páginas de Wikipedia.

Una vez determinada la lista de comunidades, utilizamos la Clasificación UNESCO (Nomenclatura Internacional de la Unesco para los campos de Ciencia y Tecnología), ‘un sistema de clasificación del conocimiento ampliamente usado en la ordenación de proyectos de investigación y de las tesis doctorales’⁴, para detectar a qué ámbitos de conocimiento podía asociarse las comunidades de artículos creadas. Las 458 comunidades fueron analizadas “manualmente” para eliminar repeticiones y dejar fuera comunidades con contenido mayoritariamente ajeno a ciencia y tecnología.

Esta tarea nos permitió eliminar 118 comunidades y establecer un **Corpus de artículos de ciencia y tecnología de 340 comunidades que engloban un total de 60.108 páginas** de la Wikipedia española. Esto supone un 5% de todos los artículos, aunque, debido a las características restrictivas de nuestro procedimiento es posible que el corpus total de artículos de ciencia y tecnología se acerque al 10% del total, es decir, alrededor de 100.000 artículos.

En el Corpus hay representados 22 de los 24 campos delimitados por la Clasificación UNESCO de dos dígitos. Como se puede observar en la tabla que sigue a continuación, no hay artículos de ‘Pedagogía’ ni de ‘Ciencias de las Artes y las Letras’. De los 22 campos representados, destaca claramente el campo de Ciencias de la tecnología que, con casi 14.000 artículos, aglutina más del 23% del conjunto de artículos del Corpus construido. En menor medida, destacan también los ámbitos de ‘Ciencia de la vida’ (14’3%) y ‘Astronomía y astrofísica’ (13’7%). Por lo tanto, los tres ámbitos con mayor peso relativo suman más de la mitad del contenido del Corpus.

³ <http://www.mapequation.org/code.html>

⁴ https://es.wikipedia.org/wiki/Clasificación_Unesco

Tabla 1. Distribución del Corpus según la Clasificación Unesco para los campos de Ciencia y Tecnología

<i>Categoría 2d UNESCO</i>	<i>Corpus</i>	
	<i>Artículos</i>	<i>%</i>
Lógica	336	0,6%
Matemáticas	3.829	6,4%
Astronomía y astrofísica	8.253	13,7%
Física	4.329	7,2%
Química	2.425	4,0%
Ciencias de la vida	8.598	14,3%
Ciencias de la tierra y el espacio	5.386	9,0%
<i>Ciencias agronómicas</i>	985	1,6%
Ciencias médicas	1.701	2,8%
Ciencias de la tecnología	13.986	23,3%
Antropología	900	1,5%
Demografía	297	0,5%
Ciencias de la economía	1.604	2,7%
<i>Geografía</i>	77	0,1%
Historia	344	0,6%
<i>Ciencias jurídicas y derecho</i>	144	0,2%
Lingüística	928	1,5%
<i>Pedagogía</i>	0	0%
Ciencias políticas	744	1,2%
Psicología	985	1,6%
<i>Ciencias de las Artes y las Letras</i>	0	0%
Sociología	2.260	3,8%
Ética	61	0,1%
Filosofía	1.936	3,2%
Total	60.108	100%

Fuente: Elaboración propia

3. Estudio de la calidad del contenido científico de la Wikipedia española

3.1. Introducción

Si ya es difícil acotar cuantitativamente el alcance del contenido científico de Wikipedia, aún lo es más evaluar su calidad. Además de la enorme cantidad de artículos que la componen, incluso si sólo consideramos el subconjunto de artículos científicos, como ya hemos visto, Wikipedia se encuentra en estado de permanente construcción y los artículos pueden estar en estadios muy diversos de elaboración: desde una simple definición de un concepto en pocas líneas, hasta los denominados artículos “destacados”, que pueden superar los 100.000 bytes de texto e incluyen material audiovisual, gráficos y tablas de datos, vínculos a otros artículos, numerosas referencias y un apartado de bibliografía, entre otros elementos.

En la Wikipedia española encontramos 1 artículo destacado por cada 1.096 artículos (**1.066** artículos destacados de un total de **1.169.455** artículos⁵). Esta proporción es prácticamente la misma que encontramos en la Wikipedia inglesa, que cuenta con un artículo destacado por cada 1.070 artículos (**4.496** artículos destacados de un total de **4.853.668** artículos⁶). Si atendemos al listado de estos artículos según el tipo de contenido (que la misma comunidad wikipedista define), observamos que solo el 34% pertenecen al ámbito de la ciencia y la tecnología (365 de los 1.067 artículos destacados). En la Wikipedia inglesa este porcentaje es inferior al 30% (1.342 de los 4.496 artículos destacados).

Tabla 2. Índice de artículos destacados en las Wikipedias española e inglesa, según su clasificación de contenidos

	contenidos	Total
Wikipedia española		
Contenidos relativos a Ciencia y Tecnología (CyT)	Astronomía y astrofísica (31) • Biología y medicina (78) • Ciencias sociales (23) • Derecho (7) • Filosofía (3) • Física, química y matemática (27) • Historia (145) • Lenguas y lingüística (7) • Prehistoria (6) • Tecnología e informática (38)	365 (34,2%)
Otros contenidos	Alimentación (14) • Fuerzas armadas y guerra (42) • Política y gobierno (57) • Arte y arquitectura (95) • Cine, fotografía, televisión y entretenimiento (88) • Cultura y sociedad (50) • Deportes (37) • Geografía y lugares (97+15) • Literatura y teatro (77) • Medio ambiente (7) • Música (88) • Religión y mitología (34)	701 (65,8%)
Total	21	1.066
Wikipedia inglesa		
Contenidos relativos a Ciencia y Tecnología (CyT)	Biology (464) • Business, economics, and finance (78) • Chemistry and mineralogy (40) • Computing (16) • Engineering and technology (44) • Geology and geophysics (23) • Health and medicine (53) • History (243) • Language and linguistics (12) • Law (65) • Mathematics (18) • Meteorology (147) • Philosophy and psychology (12) • Physics and astronomy (127)	1.342 (29,8%)
Otros contenidos	Food and drink • Politics and government • Transport • Warfare • Art, architecture, and archaeology • Culture and society • Education •	3.154 (70,2%)

⁵ Fecha de consulta, 7 de abril de 2015.

⁶ Fecha de consulta, 7 de abril de 2015.

	Geography and places • Heraldry, honors, and vexillology • Literature and theatre • Media • Music • Religion, mysticism and mythology • Royalty and nobility • Sport and recreation • Video gaming	
Total	30	1.342

Fuente: Elaboración propia

Esta observación nos permite inferir un buen nivel de calidad en los artículos de ciencia y tecnología en la Wikipedia española, puesto que mientras los artículos científicos son menos del 10% del conjunto de la Wikipedia, entre los artículos destacados la proporción supera el 30%.

Otro espacio de contenidos donde estudiar la calidad de los artículos son los denominados portales⁷, que son unas páginas de Wikipedia donde se agrupan artículos de un mismo tema. Los portales son creados por iniciativa propia de un usuario y su configuración (contenidos, apariencia, apartados...) dependen también de dicho usuario. Suelen ir relacionados con un wikiproyecto⁸ (una página donde distintos usuarios se coordinan para mejorar artículos de un mismo tema). La configuración, contenidos, apariencia, etc.

En la Wikipedia española hay un total de 432 portales⁹. Debido a la complejidad en la organización y clasificación de contenidos de Wikipedia, y que ya hemos mencionado, se hace difícil obtener una panorámica clara de la organización temática de los portales. Más allá de los portales que puedan haber quedado sin clasificar, existen cuatro listados distintos de portales en la Wikipedia española. Para nuestra pequeña observación hemos optado por tomar como referencia la página de presentación de los portales, la portada de portales¹⁰.

En la portada de portales, los 432 portales se clasifican en ocho ámbitos, de los cuales cuatro podemos considerar de ciencia y tecnología¹¹:

- Ciencias humanas y sociales (29 portales)
- Ciencias naturales y exactas (25 portales)
- Tecnología (15 portales)
- Historia (23 portales)

Por lo tanto, solo 92 de los 432 portales (el 21.3%) corresponden a temas de ciencia y tecnología. Aquí cabe destacar que el ámbito con más portales es el de *Ciencias humanas y sociales*, mientras que en nuestro corpus destacan mucho más las comunidades de tecnología. Estos 4 ámbitos engloban unos 90 portales, de temas muy variados. Por ejemplo, dentro de ciencias sociales hay un portal tan genérico como “ciencias humanas y sociales”, y un portal tan específico como “Lengua Ido”.

Estas aproximaciones ponen en evidencia la dificultad para observar y medir la calidad de los artículos de Wikipedia. A pesar de ello diferentes estudios ha intentado aproximarse a la cuestión, mediante diversas

⁷ <https://es.wikipedia.org/wiki/Wikipedia:Portal>

⁸ <https://es.wikipedia.org/wiki/Wikipedia:Wikiproyectos>

⁹ Fecha de consulta 23 de febrero de 2015.

¹⁰ <https://es.wikipedia.org/wiki/Portal:Portada>

¹¹ Hay también un ámbito de “geografía”, que si bien podría considerarse de ciencia y tecnología, al revisar “manualmente” se descartó porque la mayoría de su contenido era información varia de países, a nivel internacional.

estrategias. Wedemeyer et al. (2008), por ejemplo, han realizado un estudio comparativo entre artículos sobre ciencia publicados en la Enciclopedia Británica y sus análogos en Wikipedia; Brown (2011), por su parte, ha llevado a cabo un análisis de la calidad de los artículos de ciencia política en la Wikipedia inglesa. De hecho, el trabajo pionero en comparar la calidad de los artículos sobre temas científicos entre Wikipedia y la Británica es el conocido y controvertido artículo de Giles (2005).

Además de posibles sesgos en la cobertura temática de Wikipedia, una de las preocupaciones más comúnmente asociada a la calidad de los artículos es la falta de fuentes que sustenten de manera clara la información contenida en los artículos. Nielsen (2007) realizó un estudio pionero en el análisis de las citas y referencias a artículos científicos convencionales como indicador indirecto de calidad de los artículos de Wikipedia. Dicho estudio se centró en las referencias a artículos publicados en revistas académicas y comparó las veces que se citaba cada artículo con su factor de impacto (es decir, las veces que se citaba en publicaciones académicas), constatando que había una fuerte correlación positiva entre ambos indicadores.

Más recientemente, Ford et al. (2013) concluyen que los editores de Wikipedia consideran fiables otro tipo de fuentes más allá de las publicaciones académicas. A partir del estudio de una muestra de 500 artículos, sólo el 16% de las referencias corresponden a artículos académicos. Por otro lado, en un estudio basado en el análisis de una muestra de artículos de historia de la Wikipedia inglesa, Luyt y Tan (2010) habían ya constatado el creciente peso de las publicaciones institucionales y gubernamentales como fuentes de información.

Según Halfaker y Taraborelli (2015), la cantidad y calidad de las referencias bibliográficas en Wikipedia ha mejorado considerablemente en los últimos años. Cabe recordar, en este sentido, que dos de las políticas básicas de Wikipedia son la de *verificabilidad*, que requiere que todo artículo tenga referencias externas a la Wikipedia para que el lector pueda comprobar la exactitud y neutralidad del texto, y la de *fuentes fiables*, que indica qué tipo de fuentes se consideran más apropiadas y que confiere a las revistas científicas especializadas un valor especial.

En nuestro proyecto hemos llevado a cabo tres aproximaciones al estudio de las referencias bibliográficas, en este caso de la Wikipedia en español (una versión lingüística de la que no había hasta la fecha ningún análisis de referencias realizado). Nuestro análisis se apoya en la suposición de que la calidad de los artículos tiene mucho que ver con la calidad de las fuentes de información utilizadas por sus editores para escribirlos.

3.2. Estudio cualitativo de referencias en Wikipedia

En la primera aproximación, hemos seleccionado una muestra de 20 artículos de Wikipedia que, por un lado, forman parte de nuestro corpus de artículos de ciencia y tecnología y, por otro, forman parte del *ranking* de páginas más visitadas de la Wikipedia en español¹². De estos 20 artículos, hemos analizado sus referencias a partir de una clasificación adaptada de estudios anteriores. El listado de referencias incluye únicamente aquellas que aparecen en la sección “referencias” a pie de página del artículo, que se puede crear automáticamente, a través del uso de plantillas, al insertar una nota en el texto del artículo.

La clasificación consta de diez categorías:

Tabla 3. Clasificación de las referencias bibliográficas de un artículo científico de Wikipedia

AC	Artículo en revista científica
OC	Otras publicaciones científicas
L	Libro
P	Prensa generalista
PD	Publicaciones Divulgativas
PIC	Publicaciones Instituciones Científicas
PIP	Publicaciones Instituciones Políticas y Sociales
PC	Publicaciones Corporativas
O	Otras
Sin Cat	No se ha podido clasificar en ninguna de las anteriores categorías

Fuente: Elaboración propia

Las 20 páginas contienen un total de 994 referencias, siendo la distribución muy irregular: el 25% de las páginas tienen menos de 5 referencias, mientras los cuatro artículos con más referencias aglutinan más del 63% de las referencias (628 de las 994).

Tabla 4. Análisis de referencias en una muestra de 20 artículos de la Wikipedia española

<i>Artículo</i>	<i>Visitas</i>	<i>Referencias</i>	<i>Estadísticos:</i>	
Célula eucariota	189.892	2	<i>Media</i>	49,70
Energía	227.268	3	<i>Mediana</i>	22,50
Biología	197.673	3	<i>Desviación Std.</i>	61,8
Equinoccio	248.850	4	<i>Mínimo</i>	2

¹² https://es.wikipedia.org/wiki/Wikipedia:Ranking_de_visitas_%28Principal%29. Fecha de consulta, 15 de mayo de 2015.

Aparato digestivo	187.186	4	Máximo	225
Tabla periódica de los elementos	349.934	8	Percentil 25	5
Petróleo	190.976	9	Percentil 50	22,5
Ecosistema	244.618	13	Percentil 75	70,5
Física	234.804	17		
Átomo	233.002	22		
Fotosíntesis	198.633	23		
Psicología	196.240	30		
Sistema Solar	242.039	43		
Química	209.047	45		
Célula	326.453	69		
Electricidad	192.093	71		
Contaminación	271.510	121		
Agua	399.065	122		
Ácido desoxirribonucleico	204.092	160		
Calentamiento global	211.804	225		
Total		994		

Fuente: Elaboración propia

Los *artículos en revistas científicas* representan casi el 25% de las referencias. Prácticamente el mismo peso tienen las *publicaciones de instituciones políticas y sociales* (24,1%) y los *libros* (22,1%). Estas tres categorías suman el 70% de las referencias. Si sumamos *otras publicaciones científicas* (2%) a los *artículos en revistas científicas* (24,4%), observamos que las publicaciones científicas son el tipo de referencia con mayor frecuencia (26,4%), pero suman poco más de una cuarta parte del total.

Los *artículos en revistas científicas* no tienen el mismo peso en todas las páginas. Mientras en 'Ácido desoxirribonucleico' representan el 77'5% de las referencias (porcentaje más alto), en 'Agua' y en 'Química' no llegan al 5% (porcentaje más bajo). Sin embargo, 'Ácido desoxirribonucleico' y 'Agua' son dos de los cuatro artículos con mayor número de referencias (160 y 122, respectivamente).

Por otro lado, además de en 'Ácido desoxirribonucleico', solo en el artículo 'Célula' se observa un alto porcentaje de *artículos en revistas científicas* (55,1%). En el resto de los artículos, los *artículos en revistas científicas* obtienen porcentajes (en relación al total de referencias de cada artículo) inferiores al 20%. Es decir,

solo en 2 de los 20 artículos los *artículos en revistas científicas* representan más del 20% de las referencias (mientras que para la suma de referencias de los 20 artículos representan el 24,4% del total).

Tabla 5. Resultados del análisis de referencias de la muestra de 20 artículos

<i>(Artículo en revista científica)</i>	AC	243	24,4%
<i>(Otras publicaciones científicas)</i>	OC	20	2%
<i>(Libro)</i>	L	220	22,1%
<i>(Prensa generalista)</i>	P	69	7%
<i>(Publicaciones Divulgativas)</i>	PD	125	12,6%
<i>(Publicaciones Instituciones Científicas)</i>	PIC	39	4%
<i>(Publicaciones Instituciones Políticas y Sociales)</i>	PIP	240	24,1%
<i>(Publicaciones Corporativas)</i>	PC	17	1,7%
<i>(Otras)</i>	O	8	0,8%
<i>(no se ha podido clasificar en ninguna de las anteriores categorías)</i>	Sin Cat	13	1,3%
TOTAL		994	100%

Fuente: Elaboración propia

Estos resultados van en la línea de los mostrados anteriormente por otros estudios, pero no son concluyentes y, debido al procedimiento “manual” de clasificación de las referencias, no permiten ampliar la muestra de artículos científicos a estudiar.

3.3. Presencia de referencias de revistas académicas

En nuestra segunda aproximación al estudio de las referencias, hemos analizado la presencia de las revistas académicas de mayor impacto en el conjunto de artículos de la Wikipedia en español.

Para seleccionar el conjunto de revistas de mayor impacto hemos tomado como referencia la base de datos Web of Science. Concretamente, hemos seleccionado las cinco revistas científicas con mayor factor de impacto de cada una de las 232 áreas de conocimiento de dicha base de datos (que contiene 176 categorías para el ámbito “Ciencias” y 56 categorías para el ámbito “Ciencias Sociales”), dando un total de 1.160 revistas. Estas revistas pueden considerarse una buena aproximación a las revistas científicas de más calidad en todos los ámbitos.

Para analizar su presencia en Wikipedia, hemos diseñado un programa¹³ que identifica automáticamente todas las citas o referencias a estas 1.160 revistas en los artículos de Wikipedia en español. El primer resultado remarcable de nuestro estudio es que el 73% de las revistas (850 de las 1.160) aparecen referenciadas en Wikipedia como mínimo alguna vez. Considerando el carácter no experto de la mayoría de los editores y la naturaleza altamente especializada de estas revistas, el dato es ciertamente sorprendente.

De estas 850 revistas, 311 (por tanto más de la tercera parte de la muestra analizada), aparecen citadas más de 10 veces, y 63 de ellas más de 50 veces. Revistas multidisciplinares como Nature y Science aparecen citadas 1.625 y 1.450 veces, respectivamente, y una revista más especializada como CELL (en el ámbito de la biología celular y molecular) 180 veces. De hecho, de las 232 áreas de conocimiento, según la clasificación de Web of Science, únicamente 2 aparecen sin referencias a ninguna de sus 5 revistas con mayor factor de impacto¹⁴.

Tabla 6. Resultados de la búsqueda de revistas académicas en es.W

Área	Categorías	Revistas	Artículos	N
Ciencias	176	880	214.323	15.214
Ciencias sociales	56	280	22.299	4.199
Totales	232	1.160	236.622	19.413

Fuente: Elaboración propia

En conjunto, las 1.160 revistas de la muestra (que cuentan con 236.622 artículos publicados) han obtenido 19.413 resultados, es decir, aparecen 19.413 veces en las páginas de la Wikipedia española. En relación al número de artículos publicados por las revistas de la muestra, se podría decir que por cada 12 artículos publicados en revistas académicas, aparece una mención (a una de ellas) en la Wikipedia española. Entre el área de ‘Ciencias’ esta ratio es aún mayor, de modo que harían falta 14 artículos publicados para encontrar una mención de alguna revista en la Wikipedia española.

¹³ Un script diseñado en Python que se pasó por la Wikipedia española el 13 de octubre de 2015.

¹⁴ En concreto, estos dos ámbitos son *Ingeniería Geológica* e *Ingeniería del Petróleo*.

En cambio, en el área de 'Ciencias sociales' la ratio es mucho más baja y sólo harían falta 5 artículos publicados para encontrar una mención a alguna revista en la Wikipedia española. Teniendo en cuenta que los artículos del área de 'Ciencias sociales' representan menos del 10% del total de artículos que han publicado las revistas de la muestra, podemos decir que la presencia de revistas de 'Ciencias sociales' está sobrerrepresentada en Wikipedia.

Según el número de resultados, encontramos 21 revistas con más de 100 resultados. De éstas, 3 obtuvieron más de 1.000 resultados mientras que las otras 18 obtuvieron, aproximadamente, entre 100 y 500 resultados. Solo una de estas 21 revistas es del área de 'Ciencias sociales', siendo todas las demás de 'Ciencias'. A cierta distancia, observamos un grupo de 41 revistas con menos de 100 resultados pero más de 50. En este grupo, hay 3 revistas de 'Ciencias sociales', y el resto pertenecen al ámbito 'Ciencias'. Suman un total de 2.890 resultados.

Tabla 7. Resultados de la búsqueda de revistas académicas en es.W.

Número de revistas	Áreas		Artículos publicados	Resultado de la búsqueda en es.W
	Ciencias	Ciencias sociales		
21	20	1	12.220	>100
41	38	3	13.357	100-50
243	203	40	63.399	<50 - >10
382	264	118	77.588	2-10
162	120	42	26.339	1
311	235	76	43.719	0
Total	1.160	880	236.622	19.413

Fuente: Elaboración propia

Tabla 8. Resultados de las revistas con mayor número de resultados

Área	Disciplina	Nombre revista	Artículos publicados	Resultado de la búsqueda en es.W
Ciencias sociales	Antropología	Current Anthropology	1.757	2567
Ciencias	Ciencias Multidisciplinares	Nature	862	1625
Ciencias	Ciencias Multidisciplinares	Science	828	1450
Ciencias	Ciencias Multidisciplinares	Proceedings of the National Academy of Sciences of the United States of America	3579	516
Ciencias	Medicina General e Interna	New England Journal of Medicine	353	369
Ciencias	Conservación de la Biodiversidad	Bulletin of the American Museum of Natural History	10	353
Ciencias	Ornitología	Auk	64	238
Ciencias	Medicina General e Interna	Lancet	271	231

Ciencias	Biología Evolucionaria	Systematic Biology	74	193
Ciencias	Biología Evolucionaria	Molecular Biology and Evolution	275	191
Ciencias	Biología Molecular y Bioquímica	Cell	436	180
Ciencias	Astronomía y Astrofísica	Astrophysical Journal Supplement Series	159	171
Ciencias	Biología Celular	Cell	436	166
Ciencias	Agricultura, multidisciplinar	Journal of Agricultural and Food Chemistry	1.498	150
Ciencias	Genética y Herencia	Nature Genetics	192	144
Ciencias	Paleontología	Journal of Systematic Palaeontology	39	143
Ciencias	Ornitología	Ibis	84	139
Ciencias	Química, multidisciplinar	Chemical Reviews	281	138
Ciencias	Medicina Complementaria e Integradora	Journal of Ethnopharmacology	678	129
Ciencias	Medicina General e Interna	Annals of Internal Medicine	157	111
Ciencias	Biología	PLOS Biology	187	108
Total	<i>15 disciplinas distintas</i>		12.220	9.312

Fuente: Elaboración propia

El estudio de las referencias bibliográficas, por tanto, muestra que las fuentes de información utilizadas por los editores de Wikipedia son altamente fiables – según los estándares académicos – y, de forma indirecta, refuerzan la idea de que los artículos ofrecen contenidos de buena calidad, especialmente en temas científicos y tecnológicos. Además, estos datos abalan la idoneidad de los artículos de Wikipedia como textos introductorios muy útiles que, no sólo ofrecen una versión sintética – i.e. enciclopédica – del tema en cuestión, sino que brindan al lector referencias de alta calidad para profundizar en él.

Otro aspecto relacionado con la calidad de los artículos de Wikipedia es la exhaustividad, es decir, valorar hasta qué punto el conjunto de artículos de Wikipedia cubre los aspectos más relevantes de cada ámbito científico. En nuestro análisis de la exhaustividad hemos seleccionado un conjunto de conceptos clave de diferentes disciplinas científicas y hemos comprobado si existen las entradas correspondientes en la enciclopedia (mediante herramientas de búsqueda de contenidos en páginas web y documentos).

La muestra de conceptos clave por disciplinas académicas se ha establecido con los índices de contenidos de manuales universitarios considerados de referencia en la educación universitaria de sus respectivas materias. Para buscar la concordancia con el corpus científico del proyecto, se han escogido un total de 4 disciplinas científicas siguiendo el criterio de clasificación de la UNESCO (en este caso, la clasificación de cuatro dígitos¹⁵).

¹⁵ https://es.wikipedia.org/wiki/Clasificaci%C3%B3n_Unesco_de_4_d%C3%ADgitos

Tabla 9. Resumen de disciplinas escogidas para el análisis de exhaustividad, obras de referencia y conceptos clave extraídos.

<i>Código Unesco</i>	<i>Disciplina</i>	<i>Manual de referencia</i>	<i>Conceptos clave</i>
1202	Análisis y análisis funcional	Spivak, M. (1992). <i>Cálculo infinitesimal</i> . 2ª Edición. Ed. Reverté.	37
5307	Teoría económica	Samuelson, P.A. & Nordhaus, D. W. (2005). <i>Economía</i> (18ª edición). McGraw-Hill Interamericana.	48
5101	Antropología cultural	Harris, M. (2005). <i>Antropología cultural</i> (4ª edición). Alianza Ed.	112
2211	Física del estado sólido	Kittel, C. (1995). <i>Introducción a la física del estado sólido</i> . Ed. Reverté.	45

Fuente: Elaboración propia

Los resultados del análisis indican que la amplia mayoría de conceptos clave extraídos de los manuales académicos aparecen como artículos en la Wikipedia española, lo cual es un indicio de que la cobertura temática de la Wikipedia española en temas científicos tiende a ser exhaustiva. Los porcentajes van desde casi un 70% en el caso de los conceptos de ‘antropología cultural’ (aparecen en Wikipedia 78 de los 112 conceptos clave), hasta el 91,7% en el caso de ‘teoría económica’ (aparecen 44 de los 48 conceptos).

Tabla 10. Resumen del análisis de exhaustividad de los conceptos científicos clave, por disciplinas científicas.

<i>Disciplina</i>	<i>Conceptos</i>	<i>¿Aparecen en es.wikipedia.org?</i>	<i>¿Aparecen en nuestro corpus?</i>
Análisis y análisis funcional	Sí	32 (86,5%)	16 (43,2%)
	No	5 (13,5%)	21 (56,8%)
	Total	37 (100%)	37 (100%)
Teoría económica	Sí	44 (91,7%)	32 (66,7%)
	No	4 (8,3%)	16 (33,3%)
	Total	48 (100%)	48 (100%)
Antropología cultural	Sí	78 (69,6%)	39 (34,8%)
	No	34 (30,4%)	73 (65,2%)
	Total	112 (100%)	112 (100%)
Física del estado sólido	Sí	40 (88,9%)	38 (84,4%)
	No	5 (11,1%)	7 (15,6%)
	Total	45 (100%)	45 (100%)

Fuente: Elaboración propia

No se da la misma situación si comparamos la muestra de conceptos clave con el conjunto de artículos que conforman nuestro corpus de ciencia y tecnología (recordamos, formado por más de 60.000 artículos). En dos de las cuatro disciplinas seleccionadas, los porcentajes de concordancia son sensiblemente inferiores. En el caso de 'antropología cultural' aparecen menos del 35% de los conceptos clave. Y en el ámbito del 'análisis y análisis funcional', poco más del 43%. Este resultado nos indica que el proceso de creación del corpus de artículos de ciencia y tecnología ha sido demasiado restrictivo y ha dejado fuera algunos de los artículos que, por su temática y contenido, deberían formar parte de dicho corpus; la ventaja, para compensar esta deficiencia, es que el corpus parece estar compuesto casi exclusivamente por artículos de ciencia y tecnología, lo cual abala la fiabilidad de la nuestros análisis posteriores.

3.4. Estudio cuantitativo de referencias en Wikipedia

La utilización de citas es un elemento idóneo para explorar la relación entre las prácticas académicas tradicionales y la producción de conocimiento tal como se lleva a cabo en la Wikipedia. Es sabido los textos académicos contienen más citas y referencias que otros tipos de textos. Los científicos examinan habitualmente las referencias cuando valoran el trabajo de sus colegas con el fin de determinar si las afirmaciones que hacen están apoyadas por pruebas suficientes.

En este estudio nos hemos formulado las siguientes preguntas sobre el corpus de artículos científico-tecnológicos de Wikipedia:

1. ¿Cuáles son las revistas más citadas?
2. ¿Cuáles son los libros citados con mayor frecuencia?
3. ¿Cuál es la relación entre los libros referenciados publicados por editoriales académicas y por otros editores?

Siguiendo el espíritu de las humanidades digitales hemos utilizado programas informáticos para ayudar a responder a estas preguntas. A continuación se presentan los resultados iniciales centrados en la primera pregunta. Como mostramos más adelante, los resultados se pueden utilizar para formular nuevas preguntas. Sin embargo, en primer lugar debemos explicar cómo hemos intentado contestar la pregunta de investigación o, en otras palabras, qué metodología hemos seguido.

Metodología

Halfaker y Taraborelli (2015) son autores de un pequeño script que extrae referencias científicas de Mediawiki en caso de que contengan un identificador único, como el ISBN y el DOI, o el PMID, PMC y los identificadores utilizados en arXiv u otros repositorios abiertos de artículos académicos. Hemos utilizado esta herramienta para extraer referencias de nuestro corpus. Utilizando estos datos como punto de partida, hemos desarrollado otra secuencia de comandos que permite extraer conclusiones más precisas.

Nuestra herramienta se compone de tres partes. La primera resuelve los identificadores únicos de las referencias a libros y revistas en Wikipedia y también identifica, para los libros, la editorial. La segunda lleva a cabo diversos tipos de análisis en el conjunto de datos resultante. La tercera genera una visualización simple de los datos analizados.

El código se ha publicado bajo una licencia libre, por lo tanto, los resultados pueden ser reproducibles por parte de terceros. A pesar de ello, hemos tomado una serie de decisiones previas que vale la pena dar a conocer desde un punto de vista metodológico. En primer lugar hemos examinado y validado los ISBN. A continuación, hemos eliminado los incorrectos suponiendo que fueron recogidos así como consecuencia de errores del script anterior o por ser introducidos erróneamente por los editores humanos. Además hemos comparado los títulos

para encontrar aquellos idénticos – sin tener en cuenta mayúsculas y minúsculas – y eliminando todo lo que no eran caracteres en el alfabeto Inglés, excepto los espacios.

Los títulos de revistas están a menudo, pero no siempre, en la forma abreviada estándar. Con el fin de compararlos con eficacia era necesario resolver este problema. La norma ISO 4 define abreviaturas comunes de revistas, que parecían funcionar bien a excepción de las citas identificadas mediante PMID. Hemos abordado este problema mediante la inclusión de una segunda lista de abreviaturas que incluye todas las revistas en PMED, el repositorio de código abierto que está utilizando PMID como identificador. Ambas listas se basan en los datos proporcionados por las agencias gubernamentales de Estados Unidos.

Por último, hemos considerado que un libro era de tipo científico/académico si estaba incluido en la lista de editoriales de la Master Book List de la Web of Science (Thompson / Reuters).

Descargamos alrededor de 3 millones de páginas de Internet para recopilar la información del título y editor de cada referencia. El número es alto, porque si el programa no encuentra un identificador dado en un sitio web, lo busca en sitios web alternativos. Nuestra experiencia es que, sobre todo con los ISBN, no todos los libros se incluyen en todas las bases de datos. Por ejemplo, los siguientes servicios se comprueban para los títulos y editores de las referencias identificadas a través de ISBN:

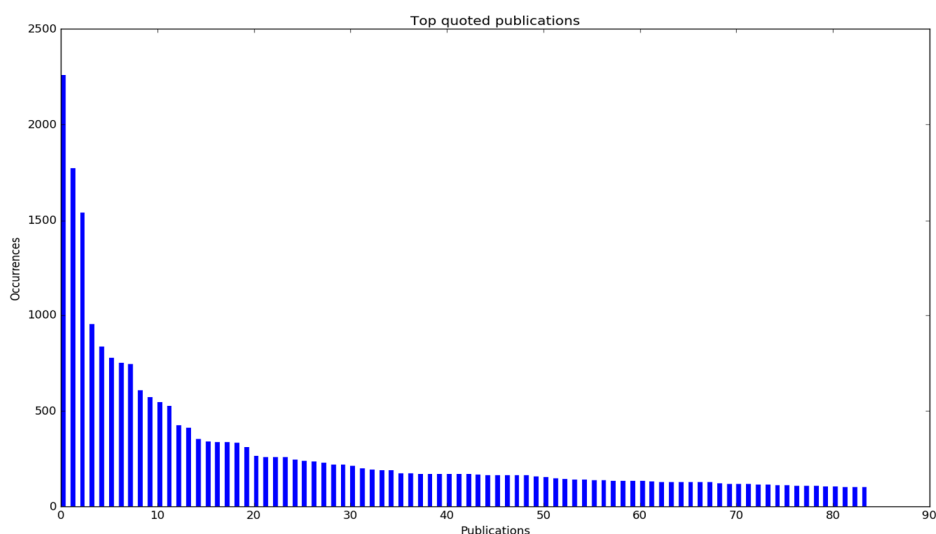
1. World Catalog
2. Bookfinder
3. Isbnsearch
4. Open Library
5. Amazon

Para hacer una sola exploración de los cerca de medio millón de referencias en nuestro corpus, el programa tarda alrededor de dos semanas en una máquina DualCore, y a pesar de que las peticiones de páginas web se han paralelizado. La tasa de éxito es de alrededor de 80% en función de la pregunta que hacemos (del tipo de análisis). Como veremos más adelante, todos estos parámetros podrían ser mejorados en versiones posteriores.

La Figura 1 muestra que la popularidad de las revistas en Wikipedia sigue una distribución de ley potencial: las tres revistas más populares quedan muy por delante de las demás, más de una docena forman una "cabeza" comparable en peso a la "cola". Por una parte, una pequeña élite de revistas - Science, The Journal of Biological Chemistry y Nature - son las que monopolizan la mayor parte de referencias. Nature aparece una segunda vez en el sexto lugar y si se suma esta cifra a la del 3er lugar, supera claramente a Science.

Sorprendentemente, la dinámica en Wikipedia es muy similar a la que puede observarse en el mundo académico. Por lo tanto, se podría argumentar que las prácticas de cita y referencia en Wikipedia siguen los patrones comunes en el mundo científico. A su vez, tanto Wikipedia como las referencias académicas reflejan las desigualdades mundiales en la producción de conocimiento y en la economía en general.

Fig 1. Frecuencia de las referencias citadas más de 100 veces



Fuente: Elaboración propia

Por otra parte, Anderson (2006) ha señalado la importancia de la “larga cola” (long tail) en la dinámica de las comunidades y los mercados en línea. Su tesis central es que el peso combinado de la cola en las plataformas de producción entre iguales mediadas tecnológicamente y en los mercados en línea es a menudo más grande que la cabeza y, por lo tanto, que la mayor parte del valor se concentra en la cola. En una línea similar, Aigrain (2012) afirma que la diversidad cultural es mayor en las comunidades de producción entre iguales. Por ejemplo, el consumo cultural de Torrents es más diversos que el de los servicios propietarios, como iTunes. De hecho, en el caso del corpus científico-tecnológico de Wikipedia investigado en este proyecto, podemos decir también que la cola es más importante que la cabeza, pero debería compararse este hallazgo con estudios similares sobre las prácticas académicas mediante estudios cuantitativos. Podría ser el caso de que a pesar de la similitud observada los editores de Wikipedia utilizaran una variedad más amplia de fuentes en su trabajo de edición.

Si observamos las revistas más citadas se confirman las tesis anteriores:

1. 2591 Science (New York)
2. 2258 The Journal of Biological Chemistry (U.S.A.)
3. 1774 Nature (London)
4. 1541 Proceedings of the National Academy of Sciences of the United States of America
5. 953 Proceedings of the National Academy of Sciences (U.S.A.)
6. 837 Nature;
7. 776 The Astrophysical Journal
8. 752 Molecular and Cellular Biology
9. 744 PLOS One
10. 607 Journal of the American Chemical Society
11. 572 Journal of Vertebrate Paleontology

Es interesante destacar que no parece haber un sesgo hacia las revistas de acceso abierto, a pesar de que algunos los modelos más refinados sí que lo han encontrado aunque sólo de forma ligera (Teplitskiy, Lu, y Duede

2015). Una pregunta que cabe formularse es ¿cómo pueden los editores tener acceso a publicaciones científicas que no son de acceso abierto y requieren, por tanto, suscripción o directamente la compra del artículo? ¿Son quizás científicos o académicos profesionales y, por tanto, tienen acceso institucional a estas fuentes? ¿O recurren al “mercado negro” y utilizan repositorios como el Science Hub Ciencia para adquirir estos artículos? De hecho, algunos estudios anteriores demuestran que su nivel educativo es, en general, alto aunque no necesariamente editan artículos en su campo profesional (Teplitskiy, Lu, y Duede 2015). Es posible, por lo tanto, que tengan acceso institucional a los repositorios sin ser “expertos” en los temas específicos en los que están trabajando en la Wikipedia.

De hecho, en una polémica decisión, la Fundación Wikimedia aceptó recientemente donaciones de cuentas, por parte de las grandes editoriales académicas como Elsevier, ScienceDirect, etc., para los editores más. Aunque esto parece sugerir que el acceso a los artículos es un verdadero problema para los editores, a la luz de los resultados de nuestra investigación, no nos lo parece.

Implicaciones

a) Evaluación

Nuestra investigación sobre el contenido científico de Wikipedia ha utilizado las referencias a las publicaciones académicas y no académicas como un indicador para abordar diversos problemas. Hemos visto que el análisis de las referencias sugiere que el contenido científico de Wikipedia no funciona, en este aspecto, de manera muy diferente a la de la producción académica convencional. Este hallazgo - interpretado teniendo en cuenta la literatura existente - sugiere la hipótesis de que los editores de Wikipedia tienen un alto nivel educativo y están familiarizados con las prácticas de producción académica – y no sólo con la forma de producción entre iguales típica de Wikipedia. Una tesis similar salió del trabajo de campo realizado por los miembros de nuestro grupo de investigación, en otro proyecto de investigación, en el terreno de los hackers y las comunidades de biohacking, a saber, que los participantes a menudo tienen experiencia profesional en un campo adyacente, de modo que son capaces de transferir sus conocimientos a sus actividades de ciencia ciudadana. Esta es otra lección que recuerda a académicos, políticos y el público en general, la necesidad de no tomar el trabajo de los “aficionados” a la ligera - ni siquiera en el campo de la comunicación pública de la ciencia.

b) Líneas de trabajo futuras

Dada la publicación en abierto del código fuente y del conjunto de datos generados, los resultados de nuestra investigación pueden ser reproducidos fácilmente. El software escrito para esta investigación se podría desarrollar, en el futuro, para obtener un conjunto de herramientas más generales para trabajar con citas y referencias en Wikipedia. Eventualmente, podría ayudar a los editores de Wikipedia a identificar referencias correctas, o convertirlo en un servicio automatizado para mejorar la calidad de las referencias académicas en la Enciclopedia libre. Ello podría contribuir, por lo tanto, a mejorar tanto la calidad científica de los contenidos como la reputación de la Wikipedia a ojos de los académicos o científicos.

4. Estudio de la presencia de la ciencia española en Wikipedia

Un elemento relevante en nuestra investigación es analizar la representatividad de Wikipedia en español como instrumento de difusión científica. En la medida en que Wikipedia puede ser considerada como una interface entre ciencia y sociedad es interesante explorar si existe una relación entre los artículos científicos en Wikipedia y los temas de mayor interés y repercusión en el ámbito de la ciencia y la tecnología en los últimos años. De ese modo, trataríamos de inferir si las páginas con contenido científico de uno de los recursos de Internet prioritarios y más utilizados para obtener información científica tienen relación con los temas actuales de investigación y los avances científicos recientes.

A fin de examinar si el contenido científico de Wikipedia se corresponde con la actividad científica desarrollada en España, hemos llevado a cabo distintos análisis. En primer lugar, se ha analizado la posible relación entre la composición del Corpus y la distribución de estudiantes matriculados en las universidades españolas. Alternativamente, se ha comparado directamente el contenido del Corpus con el conjunto de conocimientos científicos producidos y acumulados en España. Por último, se ha estudiado la presencia de las principales instituciones de investigación españolas y de los principales investigadores en las páginas de la Wikipedia española.

4.1. Estudiantes matriculados

Para analizar la posible relación entre la composición del Corpus y la distribución de estudiantes matriculados en las universidades españolas, se ha trabajado con la información estadística proporcionada por el Ministerio de Educación, Cultura y Deporte correspondiente al curso académico 2013/2014. Estos datos organizan los estudiantes en función del estudio o enseñanza que cursan y la correspondiente rama de conocimiento a la cual están adscritos, dando lugar a un total de 5 ramas de conocimiento, 8 ámbitos de estudio generales y 70 específicos. En cambio, el Corpus del proyecto utiliza la clasificación de la UNESCO, desglosada a dos dígitos, que, como hemos visto anteriormente, da lugar a 24 categorías.

Para poder realizar la comparación, hemos reagrupado los datos sobre estudiantes matriculados para que coincidan con las categorías de la clasificación UNESCO. Según esta nueva distribución, los ámbitos de estudio con mayor proporción de estudiantes matriculados son las 'ciencias tecnológicas' y las 'ciencias económicas', con porcentajes cercanos al 17% del total de estudiantes, en cada caso. En menor medida, destacan también las áreas de 'pedagogía' (11,9%) y 'medicina' (10,6%). Los datos nos muestran como la asociación es inexistente o, en el mejor de los casos, poco relevante. Es decir, los estudiantes que acceden a las instituciones de educación superior españolas no se distribuyen de acuerdo con la producción de contenido científico.

Tabla 11. Comparativa de la distribución de estudiantes universitarios con el contenido del Corpus

	<i>Artículos</i>		<i>Comunidades</i>		<i>Estudiantes</i>
Matemáticas	4.461	7,42%	46	13,53%	4,31%
Física	12.582	20,93%	27	7,94%	0,64%
Química	2.425	4,03%	17	5,00%	1,06%
Ciencias de la vida	8.266	13,75%	41	12,06%	1,76%
Ciencias de la Tierra y del espacio	5.386	8,96%	23	6,76%	1,22%
Ciencias agronómicas	1.035	1,72%	19	5,59%	1,78%
Ciencias médicas	1.651	2,75%	20	5,88%	10,62%
Ciencias de la tecnología	14.023	23,33%	78	22,94%	14,01%
Ciencias económicas	1.972	3,28%	11	3,24%	16,56%
Historia	344	0,57%	1	0,29%	2,20%
Ciencias jurídicas	144	0,24%	1	0,29%	8,04%
Lingüística	802	1,33%	8	2,35%	3,85%
Pedagogía	266	0,44%	1	0,29%	11,87%
Ciencia política	744	1,24%	6	1,76%	0,90%
Psicología	985	1,64%	14	4,12%	6,64%
Ciencias de las artes y las letras	-	0,00%	-	0,00%	5,72%
Sociología	3.062	5,09%	24	7,06%	4,89%
Ética y filosofía	1.960	3,26%	3	0,88%	0,56%
Otros	-	0,00%	-	0,00%	3,35%
	60.108	100,00%	340	100,00%	
<i>Correlación lineal</i>		<i>0,040</i>		<i>0,194</i>	

Fuente: Elaboración propia

Diferentes variables influyen lógicamente sobre el proceso de asignación de los estudiantes entre disciplinas científicas, que no son ajenas ni a las condiciones y distribución de la oferta y los distintos criterios de acceso a cada uno de los estudios (por ejemplo, las notas de corte) como también otras variables de carácter cultural o social.

Así mismo, uno de los elementos probablemente más decisivos sea el *mismatch* o desencuentro existente entre las preferencias de los estudiantes en su elección de estudio y las características de los lugares de trabajo ofrecidos en un mercado laboral afectado por un contexto de severa crisis económica y con elevado nivel de desempleo juvenil.

De ese modo, la evolución de la oferta de nuevos lugares de trabajo, sus características y la composición de la estructura productiva sean probablemente señales o indicadores más poderosos para un estudiante universitario que el nivel de excelencia o la especialidad de sus principales instituciones científicas.

4.2. Producción científica española

Para la comparación del contenido del Corpus con el conjunto de conocimientos científicos producidos y acumulados en España, hemos tomado como referencia de la producción científica española los resultados publicados en las principales revistas académicas. La muestra se ha tomado, por una lado, de la *ISI Web of Knowledge*, propiedad de Thomson Reuters y que integra a su vez tres grandes bases de datos de distintas ramas de conocimiento (*Science Citation Index (SCI)*, *Social Sciences Citation Index (SSCI)* y *Arts & Humanities Citation Index (A&HCI)*). Por el otro lado, la base de datos *Scopus*, propiedad de Elsevier y que integra resúmenes y citas de artículos en revistas científicas, libros y ponencias en conferencias.

En ambas bases de datos, se ha escogido únicamente los artículos científicos. Es decir, no se han incluido en la muestra libros o capítulos de libros, *proceedings* de conferencias, informes, notas, reseñas o cualquier otro tipo de publicación que no sea artículo científico. A su vez, ambas bases de datos disponen de un sistema de agrupación de los artículos en función de la disciplina científica a la que corresponden. Estas clasificaciones se han agrupado en los códigos de la clasificación UNESCO a dos dígitos, a fin de poder hacer la comparación con el contenido del Corpus. En el caso de *ISI Web of Knowledge* se ha limitado la muestra a los artículos de las 100 disciplinas de conocimiento con mayor nivel de producción científica, debido a la gran dimensión de la base de datos.

Tabla 12. Comparativa de la estructura de Scopus con el Corpus

	Artículos		Comunidades		Scopus
Matemáticas	4.461	7,42%	46	13,53%	3,15%
Física	12.582	20,93%	27	7,94%	8,33%
Química	2.425	4,03%	17	5,00%	6,38%
Ciencias de la vida	8.266	13,75%	41	12,06%	13,15%
Ciencias de la Tierra y del espacio	5.386	8,96%	23	6,76%	5,98%
Ciencias agrónomicas	1.035	1,72%	19	5,59%	5,15%
Ciencias médicas	1.651	2,75%	20	5,88%	28,31%
Ciencias de la tecnología	14.023	23,33%	78	22,94%	20,05%
Ciencias económicas	1.972	3,28%	11	3,24%	2,42%
Historia	344	0,57%	1	0,29%	0,00%
Ciencias jurídicas	144	0,24%	1	0,29%	0,00%
Lingüística	802	1,33%	8	2,35%	0,00%
Pedagogía	266	0,44%	1	0,29%	0,00%
Ciencia política	744	1,24%	6	1,76%	0,00%
Psicología	985	1,64%	14	4,12%	1,67%
Ciencias de las artes y las letras	-	0,00%	-	0,00%	0,94%
Sociología	3.062	5,09%	24	7,06%	0,00%
Ética y filosofía	1.960	3,26%	3	0,88%	0,00%

Otros	-	0,00%	-	0,00%	4,48%
	60.108	100,00%	340	100,00%	
<i>Correlación lineal</i>		0,519		0,598	

Fuente: Elaboración propia

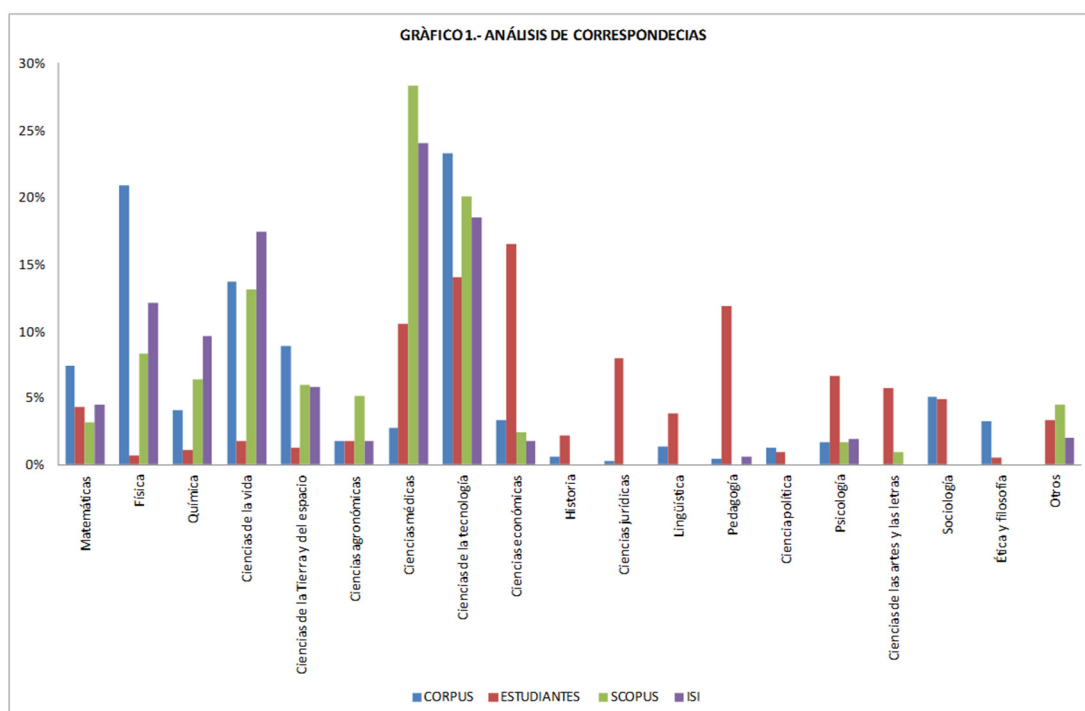
Tabla 13. Comparativa de la estructura de ISI con el Corpus

	<i>Artículos</i>		<i>Comunidades</i>		<i>ISI</i>
Matemáticas	4.461	7,42%	46	13,53%	4,49%
Física	12.582	20,93%	27	7,94%	12,16%
Química	2.425	4,03%	17	5,00%	9,64%
Ciencias de la vida	8.266	13,75%	41	12,06%	17,40%
Ciencias de la Tierra y del espacio	5.386	8,96%	23	6,76%	5,80%
Ciencias agronómicas	1.035	1,72%	19	5,59%	1,73%
Ciencias médicas	1.651	2,75%	20	5,88%	24,07%
Ciencias de la tecnología	14.023	23,33%	78	22,94%	18,48%
Ciencias económicas	1.972	3,28%	11	3,24%	1,74%
Historia	344	0,57%	1	0,29%	0,00%
Ciencias jurídicas	144	0,24%	1	0,29%	0,00%
Lingüística	802	1,33%	8	2,35%	0,00%
Pedagogía	266	0,44%	1	0,29%	0,58%
Ciencia política	744	1,24%	6	1,76%	0,00%
Psicología	985	1,64%	14	4,12%	1,93%
Ciencias de las artes y las letras	-	0,00%	-	0,00%	0,00%
Sociología	3.062	5,09%	24	7,06%	0,00%
Ética y filosofía	1.960	3,26%	3	0,88%	0,00%
Otros	-	0,00%	-	0,00%	1,99%
	60.108	100,00%	340	100,00%	
<i>Correlación lineal</i>		0,646			0,657

Fuente: Elaboración propia

Los resultados confirman, en ambos casos, la existencia de una asociación importante entre el contenido del corpus científico de Wikipedia en español y el conjunto de la producción científica mundial acumulada en el tiempo, tanto en términos de artículos como de comunidades. Esta correspondencia es todavía sensiblemente más importante en el caso de la base de datos *Web of Knowledge*.

Figura 2. Análisis de correspondencias entre Scopus, ISI y el Corpus



Fuente: Elaboración propia

Si, además, observamos en las bases de datos solamente aquella producción científica desarrollada por instituciones de investigación localizadas en España, podemos inferir que estamos haciendo una aproximación a la correspondencia entre las áreas de excelencia de la ciencia española y las áreas científicas más desarrolladas en la Wikipedia española, entendiendo como tales las que presentan mayor cantidad de artículos.

Los resultados, que se pueden ver en la tabla que sigue a continuación, muestran que la asociación es incluso ligeramente más estrecha que en la comparación anterior, e incluso mejor cuando se compara con la producción científica más reciente. Por lo tanto, la Wikipedia española reproduciría de forma relevante los nuevos desarrollos científicos en los distintos ámbitos de conocimiento existentes.

Tabla 14. Artículos científicos de instituciones españolas en Scopus y el Corpus. Comparativa.

	Artículos	Comunidades	Scopus
Matemáticas	7,42%	13,53%	4,64%
Física	20,93%	7,94%	8,41%
Química	4,03%	5,00%	8,26%
Ciencias de la vida	13,75%	12,06%	13,21%
Ciencias de la Tierra y del espacio	8,96%	6,76%	8,69%
Ciencias agrónomas	1,72%	5,59%	7,88%
Ciencias médicas	2,75%	5,88%	24,76%
Ciencias de la tecnología	23,33%	22,94%	16,07%
Ciencias económicas	3,28%	3,24%	1,90%

Historia	0,57%	0,29%	0,00%
Ciencias jurídicas	0,24%	0,29%	0,00%
Lingüística	1,33%	2,35%	0,00%
Pedagogía	0,44%	0,29%	0,00%
Ciencia política	1,24%	1,76%	0,00%
Psicología	1,64%	4,12%	1,37%
Ciencias de las artes y las letras	0,00%	0,00%	1,27%
Sociología	5,09%	7,06%	0,00%
Ética y filosofía	3,26%	0,88%	0,00%
Otros	0,00%	0,00%	3,54%
<i>Correlación lineal</i>	0,521	0,603	

Fuente: Elaboración propia

Tabla 15. Artículos científicos de instituciones españolas en WoK y el Corpus. Comparativa.

	<i>Artículos</i>	<i>Comunidades</i>	<i>WoS</i>
Matemáticas	7,42%	13,53%	3,57%
Física	20,93%	7,94%	8,72%
Química	4,03%	5,00%	10,40%
Ciencias de la vida	13,75%	12,06%	13,19%
Ciencias de la Tierra y del espacio	8,96%	6,76%	10,30%
Ciencias agronómicas	1,72%	5,59%	2,57%
Ciencias médicas	2,75%	5,88%	15,18%
Ciencias de la tecnología	23,33%	22,94%	25,82%
Ciencias económicas	3,28%	3,24%	2,65%
Historia	0,57%	0,29%	0,49%
Ciencias jurídicas	0,24%	0,29%	0,49%
Lingüística	1,33%	2,35%	0,59%
Pedagogía	0,44%	0,29%	1,04%
Ciencia política	1,24%	1,76%	0,09%
Psicología	1,64%	4,12%	2,48%
Ciencias de las artes y las letras	0,00%	0,00%	0,73%
Sociología	5,09%	7,06%	0,64%
Ética y filosofía	3,26%	0,88%	0,42%
Otros	0,00%	0,00%	0,61%
<i>Correlación lineal</i>	7,42%	13,53%	3,57%

Fuente: Elaboración propia

Finalmente, se ha procedido a examinar además la idiosincrasia de comportamiento de las instituciones de investigación españolas. En particular, se ha examinado su producción científica en el contexto internacional, a partir de la información recogida en la base de datos Web of Knowledge y teniendo en cuenta también su dinámica de evolución. Los resultados nos muestran una mejora evidente de la participación de la producción científica elaborada en España. En los últimos cinco años se ha publicado más del 21% de todos los artículos incluidos en la base de datos.

Tabla 16. Evolución de la presencia de la producción científica española. 2011-2015

<i>Período</i>	<i>Web of Science</i>	<i>Instituciones españolas</i>	<i>Participación</i>
1980 y anteriores	2.466.316	1.731	0,07%
1981-1990	3.088.797	12.349	0,40%
1991-2000	10.526.175	195.836	1,86%
2001-2010	20.569.577	527.839	2,57%
2011-2015	9.869.976	237.141	2,40%
Conjunto del período	46.520.841	974.896	2,10%

Fuente: Elaboración propia

Si nos centramos en las publicaciones correspondientes al período 2011-2015, observamos que las instituciones de investigación españolas presentan sus principales ámbitos de excelencia en la comunicación audiovisual, la astronomía y la astrofísica, la geografía, la biología marina, las ciencias de la alimentación, la gestión de recursos energéticos, la geología y algunos ámbitos de las humanidades, como la lingüística y la literatura clásica.

Sin embargo, en los ámbitos de mayor producción científica incluidos en la base de datos ISI Web of Knowledge la presencia de las publicaciones españolas es muy inferior, con excepción de las disciplinas de las ingenierías, la química y la física.

Tabla 17. Presencia en las áreas de excelencia ISI.

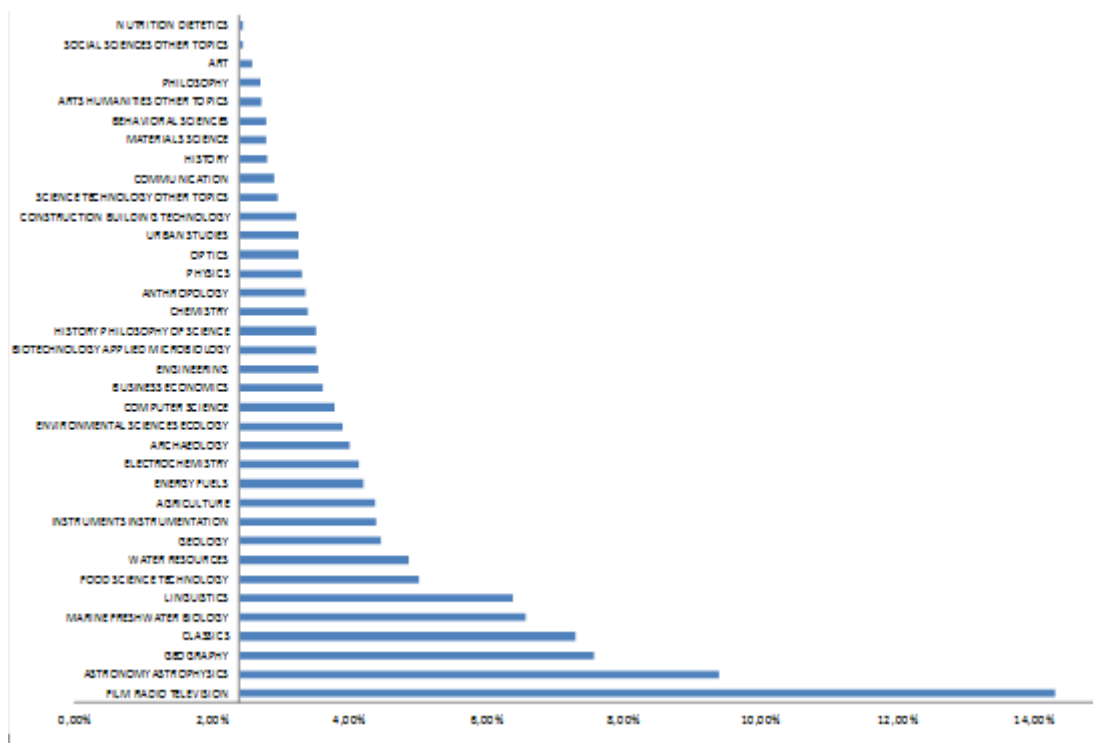
	<i>Conjunto base ISI WoS</i>	<i>Producción científica española</i>	<i>Participación</i>
Biología y bioquímica molecular	1.392.871	9.852	0,71%
Farmacología y farmacia	872.710	5.547	0,64%
Química	787.067	26.796	3,40%
Herencia genética	724.827	3.818	0,53%
Ingeniería	676.869	24.000	3,55%
Física	634.845	21.024	3,31%
Matemáticas	605.385	10.066	1,66%
Biología celular	567.343	3.611	0,64%
Neurociencia	550.127	7.881	1,43%
Psicología	532.713	6.446	1,21%

ISI Web of Knowledge	9.869.976	237.141	2,40%
----------------------	-----------	---------	-------

Fuente: Elaboración propia

La presencia internacional de la ciencia española está mermada además por un déficit de carácter distributivo. En este sentido, si se seleccionan las 100 disciplinas con mayor producción científica a escala global y se calcula la participación de las instituciones españolas en los primeros decimales se observa como su peso es inferior a la media del conjunto de la base de datos.

Figura 3. Presencia de la producción científica de instituciones españolas 2011-2015 (% Publicaciones en *Web of Science* por disciplina académica)



Fuente: Elaboración propia

De ese modo, si bien es evidente una creciente implicación de las instituciones científicas españolas en la publicación de artículos de referencia, la circunstancia de que la producción científica se oriente hacia ámbitos de disciplinas con menor actividad relativa es un sesgo de carácter asignativo que dificulta el progreso de la diseminación de las actividades científicas desarrolladas en España.

4.3. Presencia de instituciones científicas e investigadores

El tercer análisis de la presencia de la ciencia española en Wikipedia consiste en una búsqueda (mediante herramientas de búsqueda de contenidos en páginas web) de las principales instituciones de investigación españolas y de los principales investigadores en las páginas de la Wikipedia española. Ante la complejidad de establecer cuál podría ser el listado de instituciones científicas e investigadores más relevantes del panorama científico español, se optó por usar como referencia algún *ranking* reconocido internacionalmente, sin que eso suponga que consideremos que sea la mejor manera de valorar las instituciones científicas o los investigadores.

En el caso de las instituciones científicas, se ha determinado el *ranking* a partir del SCIMAGO Institutions Ranking¹⁶, donde para España y 2014 se recogen 243 entidades (que incluyen universidades, centros biomédicos y centros de investigación). El *ranking* distingue nueve indicadores de investigación. Para confeccionar nuestra muestra hemos seleccionado las cinco primeras instituciones de cada uno de los nueve indicadores. Eliminando aquellas que aparecen entre las cinco primeras posiciones en más de un indicador, nos queda una lista de 26 entidades, compuesta por 17 centros de investigación gubernamentales, 6 universidades y 3 centros biomédicos.

En conjunto, los resultados de la búsqueda nos revelan que el CSIC es la institución científica española con mayor representación en Wikipedia, puesto que además del propio centro, otras ocho entidades de la muestra son centros adscritos o pertenecientes al CSIC. En la búsqueda, la página del CSIC obtiene más de 5.000 resultados, mientras que la suma del resto de entidades no supera los 4.000 resultados. Es decir, la entidad con mayor representación de la muestra tiene más resultados que la suma de todos los demás.

Tienen página propia – es decir, un artículo o entrada- en Wikipedia 14 de las 26 entidades de la muestra, poco más de la mitad. El hecho de tener página propia no significa tener un mayor número de resultados (aparecer más veces en la Wikipedia española). Por otro lado, de las 26 entidades de la muestra, doce tienen 10 o menos resultados; muchos de ellos, además, poco relevantes (menciones en anexos o listas de premios, por ejemplo).

Aparte del CSIC, las entidades con mayor presencia en Wikipedia son las universidades (entre 500 y 800 resultados por universidad), que cuentan con una amplia diversidad de páginas propias (facultades y centros propios, bibliotecas, proyectos, alumni...).

De las 26 instituciones de la lista, 12 tienen 10 o menos resultados en es.Wikipedia; muchos de ellos, además, poco relevantes. En general, la mayoría de resultados no son de artículos científicos sino de listados de centros, entidades y científicos.

Tabla 18. Presencia en es.W de las principales instituciones científicas españolas (SCIMAGO)

<i>Institución científica</i>	<i>¿Tiene página propia en es.W?</i>	<i>Resultados relevantes</i>
Centre de Recerca en Epidemiologia Ambiental	No	2
Centre de Recerca en Sanitat Ambiental	No	1
Centre d'Investigació en Nanociència i Nanotecnologia (CSIC)	No	3
Centro de Ciencias Humanas y Sociales (CSIC)	Sí	>100
Centro de Regulación Genómica	No	10
Centro Nacional de Investigaciones Oncológicas	Sí	>40
Consejo Superior de Investigaciones Científicas, CSIC	Sí	>5.000
Deustuko Unibertsitatea	Sí	>400
Institut Català de Recerca de l'Aigua	No	3
Institut Català d'Investigació Química	No	3
Institut de Ciències de l'Espai (CSIC)	Sí	20
Institut de Diagnosi Ambiental i Estudis de l'Aigua	No	1
Institut de Física d'Altes Energies	Sí	10
Instituto de Astrofísica de Andalucía (CSIC)	Sí	>50
Instituto de Astrofísica de Canarias	Sí	>90
Instituto de Física Corpuscular (CSIC)	Sí	>10
Instituto de Física de Cantabria (CSIC)	No	10
Instituto de la Grasa de Sevilla (CSIC)	No	4
Instituto de Tecnología Química (CSIC)	No	3
Instituto Nacional del Carbón (CSIC)	Sí	19
Universidad Autónoma de Madrid	Sí	498
Universidad Complutense de Madrid	Sí	515
Universitat Autònoma de Barcelona, UAB	Sí	>700
Universitat de Barcelona, UB	Sí	>800
Universitat de València	Sí	>600
Xarxa de Referència en Tecnologia dels Aliments	No	0

Nota: Algunos resultados no son números exactos porque al haber distintas búsquedas (por distintas denominaciones, siglas...) no se ha comprobado exhaustivamente la existencia de resultados repetidos.

Fuente: Elaboración propia

Para poder determinar hasta qué punto estos resultados permiten deducir una mayor o menor presencia en Wikipedia, hemos repetido el análisis para otros países europeos. En concreto, hemos seleccionado Holanda, Italia y Reino Unido (básicamente, porque para 2014 tenían índices similares de inversión estatal en I+D). Para estos tres países, hemos repetido el procedimiento de selección: a partir del *ranking* SCIMAGO, también con los resultados de 2014, se han seleccionado los cinco primeros centros de cada uno de los nueve indicadores, eliminando los centros que aparecen en más de un indicador.

En el caso de Holanda, la muestra se compone por un total de 9 entidades. Este ya es un dato a tener en cuenta, puesto que mientras en el caso de España la muestra (que podría ser de 45 entidades) es de 26 entidades tras eliminar los centros que aparecen en más de un indicador, en el caso de Holanda la muestra se queda en solo 9 entidades, lo que significa que hay un gran número de centros que destacan en más de un indicador de producción científica.

De los 9 centros holandeses estudiados, 7 tienen página propia en la Wikipedia holandesa (nl.W). Solo un centro no tiene presencia alguna en Wikipedia. Para los demás, las dos universidades que aparecen en la lista son las entidades que obtienen un mayor número de resultados, y se sitúan a mucha distancia del resto. Mientras la Universidad de Utrech obtiene alrededor de 1.800 resultados, y la Universidad de Amsterdam llega casi a los 3.000, el Netherlands Institute for Radio Astronomy y el FOM Institute for Atomic and Molecular Physics (AMOLF) no llegan a los 20 resultados.

Tabla 19. Presencia en nl.W de las principales instituciones científicas españolas (SCIMAGO)

<i>Institución científica</i>	<i>¿Tiene página propia en nl.W?</i>	<i>Resultados relevantes</i>
Utrecht University (UU)	Sí	50 (en inglés) 1.800 (holandés)
University of Amsterdam (UvA)	Sí	24 (en inglés) 2.990 (holandés)
National Institute for Subatomic Physics (Nikhef)	Sí	0 (inglés) 6 (holandés) 26 (acrónimo)
Materials Innovation Institute (M2i)	No	0
FOM Institute for Atomic and Molecular Physics (AMOLF)	Sí	15 (sólo inglés)
European Space Research and Technology Centre (ESTEC)	Sí	39 (sólo inglés)
Royal Netherlands Academy of Arts and Sciences (KNAW)	Sí	<1.000 (16 en inglés)
Netherlands Institute for Radio Astronomy	Sí	<20 (3 en inglés)
TI Food and Nutrition (TIFN)	No	2 (acrónimo)

Fuente: Elaboración propia

En el caso de Italia, la muestra se compone de un total de 10 entidades, de las cuales 8 tienen página propia en la Wikipedia italiana. En esta muestra, como en el caso español, destacan las instituciones públicas. Sin embargo, las dos entidades que obtienen mayor número de resultados, cerca de mil cada una, son las dos únicas universidades de la muestra. En tercer lugar aparece lo que podría considerarse el equivalente al CSIC en España, el Consiglio Nazionale delle Ricerche, si bien no supera los 500 resultados. Cabe destacar que la mitad de la muestra tiene menos de veinte resultados cada una.

Tabla 20. Presencia en it.W de las principales instituciones científicas españolas (SCIMAGO)

<i>Institución científica</i>	<i>¿Tiene página propia en it.W?</i>	<i>Resultados relevantes</i>
Consiglio Nazionale delle Ricerche * (CNR)	Sí	490
Universita degli Studi di Roma La Sapienza (Sapienza Università)	Sí	1.000 aprox.
Agenzia Spaziale Italiana (A.S.I.)	Sí	<200
Politecnico di Torino	Sí	949
Istituto di Ricerca per la Protezione Idrogeologica - CNR (IRPI)	Sí	<20
Novartis Farma, S.p.A., Italy (NOVARTISML)	Sí	<130
Institute for Health and Consumer Protection (IHCP)	Sí	7
Consorzio Nazionale Interuniversitario per le Telecomunicazioni (CNIT)	Sí	10
Istituto Gas Ionizzati - CNR (IGI/ Consorcio RFX)	No	8
Istituto Officina dei Materiali - CNR (IOM)	No	2

Fuente: Elaboración propia

En el caso de Reino Unido, la muestra se compone de un total de 9 entidades, todas ellas cuentan con página propia en la Wikipedia inglesa (en.W). Aquí los resultados son muy superiores a los vistos en los casos anteriores. Seis de las nueve entidades obtienen más de mil resultados. Las dos entidades que más destacan son universidades y se sitúan alrededor de los veinte mil resultados. Por el otro extremo, destaca el centro de astronomía UKATC, que no supera los cuarenta resultados; pero el resto de centros tienen más de cuatrocientos.

Tabla 21. Presencia en en.W de las principales instituciones científicas españolas (SCIMAGO)

<i>Institución científica</i>	<i>¿Tiene página propia en en.W?</i>	<i>Resultados relevantes</i>
University of Oxford	Sí	19.300
University of Cambridge	Sí	22.600

Wellcome Trust	Sí	2.520
School of Oriental and African Studies (SOAS)	Sí	2.670
Cancer Research UK	Sí	1.320
Centre for Economic Policy Research (CEPR)	Sí	<400
Wellcome Trust Sanger Institute	Sí	842
UK Astronomy Technology Centre (UKATC)	Sí	39
Medical Research Council (MRC)	Sí	1.880

Fuente: Elaboración propia

Comparando los resultados obtenidos para los distintos países, destaca en primer lugar que tanto en Holanda, como en Italia y Reino Unido, las entidades con mayor presencia son universidades, mientras que en España el protagonismo se lo lleva el CSIC (y sus centros adscritos).

Si atendemos únicamente a los valores absolutos de resultados, Reino Unido es el país con mayor presencia de centros de investigación en sus páginas de Wikipedia. España obtiene resultados absolutos muy por encima de Holanda e Italia, pero si calculamos el peso relativo de los resultados en función del número de entidades de las muestras, observamos que España se sitúa en segundo lugar: si bien el total de resultados es prácticamente el doble que en el caso holandés, Holanda obtiene más de 545 resultados por centro de investigación, mientras que España obtiene 342. También en términos relativos Reino Unido destaca sobre el resto de países estudiados, con más de 5.730 resultados por centro de investigación.

Tabla 22. Presencia en Wikipedia de las principales instituciones científicas españolas (SCIMAGO). Comparativa

País	Muestra	Centros con página propia en su respectiva versión lingüística de Wikipedia	Resultados tras buscar su presencia en su respectiva versión lingüística de Wikipedia	Resultados por centro de investigación	Tipo de entidades con mayor número de resultados	Posición en el ranking de Wikipedias
España	26	14 (53'8%)	8.892	342	CSIC	10
Holanda	9	7 (77'8%)	4.911	545'7	Universidades	5
Italia	10	8 (80%)	2.816	281'6	Universidades	8
Reino Unido	9	9 (100%)	51.571	5.730'1	Universidades	1

Fuente: Elaboración propia

Contrariamente a lo que cabría suponer, este orden de resultados no se corresponde con el *ranking* de Wikipedias por número total de artículos publicados¹⁷. En dicho listado, la versión inglesa se sitúa en primer lugar y la versión holandesa en quinto; la Wikipedia italiana la encontramos en octavo lugar y la española en el décimo.

En segundo lugar, hemos estudiado la presencia en Wikipedia de los investigadores españoles más destacados (también mediante herramientas de búsqueda de contenidos en páginas web). Para establecer la muestra hemos recurrido a la clasificación *Highly Cited Researchers* de la agencia Thomson Reuters¹⁸. En su página web se detalla la lista completa de investigadores (3.215), con sus nombres, categoría (área de investigación), primera afiliación y segunda afiliación, con datos actualizados a diciembre de 2014.

La búsqueda por *Spain* nos da un total de 49 resultados. 43 tienen un centro español como afiliación principal y los 6 restantes lo tienen como segunda afiliación. La muestra ha tenido en cuenta estos 43 investigadores que tienen un centro español como filiación principal. 7 de ellos cuentan con página propia en la Wikipedia española (el 15'2% de la muestra). Tan solo 3 investigadores aparecen en más de 10 páginas, siendo uno de ellos, Avelino Coma, el único que destaca (con alrededor de 50 resultados). Curiosamente, este investigador no pertenece a ningún centro del CSIC, entidad que hemos visto anteriormente que tiene una presencia destacada en Wikipedia, en relación al resto de centros de investigación españoles. 16 investigadores (34'8% de la muestra) no aparece en ninguna página de la Wikipedia española. Por otro lado, 8 investigadores (17'4%) obtienen un único resultado.

Tabla 23. Presencia en es.W de los investigadores españoles más destacados (Thomson Reuters)

<i>Científico</i>	<i>Afiliación</i>	<i>¿Tiene página propia en es.W?</i>	<i>Resultados relevantes</i>
Begoña Bartolomé	Institute of Food Science Research (CIAL-CSIC)	No	3
Jose Luis Domingo	University Rovira i Virgili	No	0
Carmen Gómez-Cordovés	CSIC	No	2
Francisco A. Tomas-Barberan	CSIC	No	6
Andy Pohl	Ctr Genom Regulat CRG	No	0
Juan Bisquert	Universitat Jaume I	No	0
Avelino Corma	Instituto de Tecnologia Quimica (UPV-CSIC)	Sí	50
Luis M. Liz-Marzan	CIC biomaGUNE	No	1
Miguel Yus	University of Alicante	No	1
Joan Bladé	Hosp Clin Barcelona	No	1

¹⁷ <https://es.wikipedia.org/wiki/Anexo:Wikipedias>

¹⁸ <http://highlycited.com>

F. Xavier Bosch	Inst Catala Oncol IDIBELL	No	0
Jordi Bruix	University of Barcelona	No	1
Elías Campo	University of Barcelona	Sí	15
Valentín Fuster	CNIC	Sí	15
Josep M. Llovet	University of Barcelona (75%)	No	0
Salvador Garcia	University of Granada	No	0
Francisco Herrera	University of Granada	No	1
David Posada	University of Vigo	No	2
Alfonso Valencia	Spanish Natl Canc Res Ctr	No	2
Jordi Gali	University of Pompeu Fabra	Sí	8
Enrique Herrera-Viedma	University of Granada	No	1
Jose Matas	UPC Universitat Politècnica de Catalunya	No	0
Miguel B. Araujo	CSIC	No	1
Jordi Bascompte	CSIC	No	5
Josep Peñuelas	Autonomous University of Barcelona	Sí	5
Montserrat Vila	Estn Biol Donana EBD, CSIC	No	0
Andrés Alastuey	CSIC	No	0
Xavier Querol	CSIC	No	0
Juan J. Nieto	Universidad de Santiago de Compostela, USC	No	0
Rosana Rodríguez-López	Universidad de Santiago de Compostela, USC	No	0
Christian Brander	Hosp Badalona Germans Trias & Pujol	No	3
Roderic Guigo	Ctr Genom Regulat CRG	No	4
José Manuel Garcia-Verdugo	University of Valencia	Sí	4
Francisco J. Garcia-Vidal	Universidad Autonoma de Madrid (UAM)	No	0
Francisco Guinea	CSIC	No	2
Maciej Lewenstein	ICFO - Institut de Ciencies Fotoniques	No	2
Luis Martín-Moreno	University of Zaragoza	No	0
Sergei D. Odintsov	ICREA and ICE(CSIC-IEEC)	Sí	3

Jaume Flexas	University of Illes Balears	No	0
Roberto Solano	Ctr Nacl Biotecnol, CSIC	No	1
Eduard Vieta	University of Barcelona	No	2
Ismael Rafols	Polytechnic University of Valencia	No	0

Fuente: Elaboración propia

La búsqueda por *Holland* en la clasificación *Highly Cited Researchers* nos da una muestra de 45 investigadores (tampoco aquí hemos tenido en cuenta los investigadores que tienen Holanda como segunda filiación). El 22'2% de los investigadores (10) tienen página propia en la Wikipedia holandesa. Los investigadores con más de 10 resultados son 7 (15'6%) y el investigador que más destaca cuenta con 78 resultados (con afiliación a un centro universitario). Hay 21 investigadores (el 46'6% de la muestra) que no aparecen en ninguna página de la Wikipedia holandesa, y otros 8 (17'8%) obtienen un único resultado.

Tabla 24. Presencia en nl.W de los investigadores holandeses más destacados (Thomson Reuters)

<i>Científico</i>	<i>Afiliación</i>	<i>¿Tiene página propia en nl.W?</i>	<i>Resultados relevantes</i>
Pedro W. Crous	CBS KNAW Fungal. Biodivers. Ctr.	Sí	1
John List	CentER	No	0
Cornelia M. Van Duijn	Ctr. Med. Syst. Biol.	No	3
Cees Dekker	Delft Univ. Technol.	Sí	78
Leo P. Kouwenhoven	Delft Univ. Technol.	No	0
Ulrich S. Schubert	DPI	No	0
E.W. Meijer	Eindhoven Univ. Technol.	Sí	28
Martijn M. Wienk	Eindhoven Univ. Technol.	No	0
Rene AJ Janssen	Eindhoven Univ. Technol.	Sí	1
Albert Hofman	Erasmus Univ. Med. Ctr.	No	6
Albert (Ab) Osterhaus	Erasmus Univ. Med. Ctr.	Sí	17
Andre G. Uitterlinden	Erasmus Univ. Med. Ctr.	No	3
Ben A. Oostra	Erasmus Univ. Med. Ctr.	No	1
Fernando Rivadeneira	Erasmus Univ. Med. Ctr.	No	2
Guus F. Rimmelzwaan	Erasmus Univ. Med. Ctr.	No	0
Henning Tiemeier	Erasmus Univ. Med. Ctr.	No	2
Jacqueline CM. Witteman	Erasmus Univ. Med. Ctr.	No	0

Johan P. Mackenbach	Erasmus Univ. Med. Ctr.	No	2
Pieter Sonneveld	Erasmus Univ. Med. Ctr.	No	0
Ron A.M. Fouchier	Erasmus Univ. Med. Ctr.	No	0
Yurii S. Aulchenko	Erasmus Univ. Med. Ctr.	No	0
Arie Pieter Kappetein	Erasmus Univ. Rotterdam	No	0
Arnold B. Bakker	Erasmus Univ. Rotterdam	Sí	10
Eric Boersma	Erasmus Univ. Rotterdam	No	0
Henk Volberda	Erasmus Univ. Rotterdam	Sí	13
Justin Jansen	Erasmus Univ. Rotterdam	No	11
Patrick W.J.C. Serruys	Erasmus Univ Rotterdam, Netherlands	No	0
Hans Clevers	Hubrecht Inst.	Sí	15
Desiree MFM van der Heijde	Leiden Univ. Med Ctr.	No	0
Jeroen J. Bax	Leiden Univ.	No	0
Martin J. Schalijs	Leiden Univ.	No	1
Wim Jiskoot	Leiden Univ.	No	0
Harry JGM Crijns	Maastricht Univ Med Ctr.	No	0
Jim van Os	Maastricht Univ.	Sí	12
Marcel E. Visser	Netherlands Inst Ecol NIOO KNAW	No	0
Ron De Graaf	Netherlands Inst Mental Hlth & Addict	No	1
Stefan Schouten	NIOZ Royal Netherlands Inst Sea Res	No	5
Christian F. Beckmann	Radboud Univ Nijmegen	No	1
Lambertus A. Kiemeny	Radboud Univ Nijmegen	No	0
Mikhail I. Katsnelson	Radboud Univ Nijmegen	Sí	9
Anton E. Kunst	Univ. Amsterdam	No	0
Harry R. Buller	Univ. Amsterdam	No	1
Hergen Spits	Univ. Amsterdam	No	0
John J.P. Kastelein	Univ. Amsterdam	No	0
Karsten Kalbitz	Univ. Amsterdam	No	1

Fuente: Elaboración propia

La búsqueda por *Italy* en la clasificación *Highly Cited Researchers* nos da una muestra de 40 investigadores. En este caso, el porcentaje de investigadores con página propia en la respectiva Wikipedia baja al 7'5% (3 investigadores). Por otro lado, el número de investigadores que no aparecen en ninguna página asciende a 21 (el 52'5%), y los que obtienen un único resultado son el 15% (6 investigadores). Con más de 10 resultados, 3 investigadores, si bien el resultado máximo observado es de 17.

Tabla 25. Presencia en it.W de los investigadores italianos más destacados (Thomson Reuters)

<i>Científico</i>	<i>Afiliación</i>	<i>¿Tiene página propia en it.W?</i>	<i>Resultados relevantes</i>
Aldo Prieto Maggioni	ANMCO Res Ctr, Italy	No	0
Riccardo Valentini	Centro Euro-Mediterraneo sui Cambiamenti Climatici, Italy	No	0
Serena Sanna	Cittadella Universitaria Cagliari, Italy	No	0
Alessandro Reali	CNR, Italy	No	0
Manuela Uda	CNR, Italy	No	0
Vicenzo Di Marzo	CNR, Italy	No	4
Frank J. Dentener	Commiss European Communities, Italy	No	0
Gianni Tognoni	Consorzio Mario Negri Sud, Italy	No	6
Antonio Colombo	Ctr Cuore Columbus, Italy	No	0
Aron Goldhirsch	European Inst Oncol, Italy	No	0
Luigi Tavazzi	Fondazione IRCCS Policlinico S. Matteo, Italy	No	1
Alberto Mantovani	Humanitas Clinical and Research Center, Italy	No	5
Gianni Zamorani	INAF Osservatorio Astron Bologna, Italy	No	0
Alvio Renzini	INAF Osservatorio Astron Padova, Italy	No	0
Sandro Fuzzi	Inst Atmospher Sci & Climate ISAC, Italy	No	0
Mauro Serafini	Ist Nazl Ric Alimenti & Nutr, Italy	No	1
Alessandra Carattoli	Ist Super Sanita	No	0
Bruno Scrosati	Italian Institute of Technology	No	1
Paolo Giommi	Italian Space Agency (ASI)	No	0
Giuseppe Remuzzi	Mario Negri Inst Pharmacol Res	Sí	13
Maria Cristina Facchini	Natl Res Council CNR	No	1
Mario Boccardo	Osped Molinette	No	2

Salvatore Siena	Osped Niguarda Ca Granda	No	4
Daniele Ielmini	Politecn Milan	No	0
Erasmus Carrera	Politecn Torino	No	0
Nicola Bellomo	Politecn Torino	No	3
Andrea Cimatti	Univ Bologna	No	0
Michele Bacarani	Univ Bologna	No	4
Stefano Volinia	Univ Ferrara	No	1
Andrea Scozzafava	Univ Florence	No	1
Paolo Nannipieri	Univ Florence	No	0
Giuseppe Mancia	Univ Milano Bicocca	Sí	17
Francesco M. Veronese	Univ Padua	No	2
Calogero Vetro	Univ Palermo	No	0
Daniele Del Rio	Univ Parma	No	0
Giuseppe Mingione	Univ Parma	Sí	13
Nicoletta Pellegrini	Univ Parma	No	2
Maurizio Battino	Univ Politecn Marche	No	2
Giovanni Petrone	Univ Salerno	No	0
Giovanni Spagnuolo	Univ Salerno	No	0

Fuente: Elaboración propia

Sorprendentemente, La búsqueda por *United Kingdom* en la clasificación *Highly Cited Researchers* nos da una muestra de solo 2 investigadores, por lo que en este apartado no tendremos en cuenta los datos de Reino Unido para la comparativa entre países.

Tabla 26. Presencia en en.W de los investigadores ingleses más destacados (Thomson Reuters)

<i>Científico</i>	<i>Afiliación</i>	<i>¿Tiene página propia en en.W?</i>	<i>Resultados relevantes</i>
Sarah E. Hunt	European Bioinformat Inst, United Kingdom	No	<20
Michael L. Dustin	The Kennedy Inst Rheumatology, United Kingdom	No	8

Fuente: Elaboración propia

La comparación de los resultados para España, Holanda e Italia nos puede sugerir distintas conclusiones, en función del indicador que consideremos más relevante. Mientras España tiene el menor porcentaje de investigadores que no aparecen en ninguna página de Wikipedia, los resultados absolutos no son tan altos como en Holanda y los investigadores con página propia son menos. En todo caso, Italia es el caso que claramente cuenta con una menor presencia de investigadores en su versión lingüística de Wikipedia, en relación al resto de países de la muestra.

Tabla 27. Presencia en Wikipedia de los investigadores más destacados (Thomson Reuters). Comparativa.

	<i>España</i>	<i>Holanda</i>	<i>Italia</i>
Investigadores de la muestra	46	45	40
Investigadores con página propia	7 (15'2%)	10 (22'2%)	3 (7'5%)
Investigadores con más de 10 resultados	3 (6'5%)	7 (15'6%)	3 (7'5%)
Resultado más alto	50	78	17
0 resultado	16 (34'8%)	21 (46'6%)	21 (52'5%)
1 resultados	8 (17'4%)	8 (17'8%)	6 (15%)

Fuente: Elaboración propia

En conclusión, esta aproximación a la presencia de centros de investigación e investigadores en Wikipedia es de carácter exploratorio. Los resultados no son concluyentes y solo nos permiten intuir que la representación de los centros de investigación y sus investigadores en Wikipedia no parece tener relación con la producción científica de dichos centros e investigadores. En comparación con otros países europeos, España no destaca por unos mayores ni unos menores resultados. Finalmente, cabe destacar que en el caso de los investigadores la mayoría de resultados encontrados eran citas de co-autoría de artículos académico, de modo que, teniendo en cuenta la baja proporción de páginas propias entre la muestra de investigadores, se intuye un mayor esfuerzo en incluir contenidos académicos como referencias en los artículos que no en dar visibilidad a los científicos (y centros de investigación) que producen tales contenidos.

5. Estudio de los editores y del proceso de edición

Uno de los hechos más sorprendentes de Wikipedia, especialmente si atendemos a su contenido científico y tecnológico, es que está construida - escrita y editada - mayoritariamente por legos, es decir, por personas no expertas. De hecho, en la larga historia de los proyectos enciclopédicos, es la primera vez que se crea una iniciativa de este tipo al margen de la comunidad científica o académica (O'Sullivan, 2011) y, curiosamente, parece la más exitosa desde varios puntos de vista.

La participación preponderante de legos en la construcción de Wikipedia no es casual, dado que Wikipedia está deliberadamente diseñada para que cualquier persona - con unos conocimientos y habilidades de escritura mínimas - pueda editarla: desde alumnos de primaria a científicos especializados; pero en ningún caso se requieren acreditaciones formales - títulos académicos, por ejemplo- para poder editar artículos. Hay que recordar, en este sentido, que Wikipedia nace a partir del fracaso de un proyecto anterior, Nupedia, que se basaba en la colaboración de expertos, como editores y revisores, con objeto de crear una enciclopedia virtual. En Wikipedia, en cambio, los artículos están escritos en su mayoría por no expertos y los expertos que contribuyen a menudo lo hacen sobre materias ajenas a su experiencia profesional (Jemielniak, 2014, 3). Esta es una de las diferencias fundamentales con las comunidades y proyectos de software libre donde el nivel de pericia requerido para participar es claramente superior.

El hecho de que los artículos de Wikipedia estén mayoritariamente escritos por legos y que la participación no dependa de ningún tipo de acreditación previa, ha sido causa, tradicionalmente, de un cierto escepticismo y/o desconfianza respecto a su calidad y fiabilidad, particularmente desde del entorno académico.

5.1. Estudio de los editores de artículos científicos de la Wikipedia española

No hay en realidad muchos datos empíricos sobre el perfil sociodemográfico de la comunidad de editores de Wikipedia. En una publicación de la Fundación Wikimedia se recogió esta inquietud y se llevó a cabo una recogida de datos entre editores de Wikipedia a nivel internacional. Según esos datos, publicados en 2011¹⁹, el perfil de editor de Wikipedia es un hombre con estudios universitarios, sobre los 30 años, usuario familiarizado con los ordenadores (*computer savvy*, en inglés) pero no necesariamente un programador, no dedica (a pesar del tópico) mucho tiempo a los videojuegos, y vive en Estados Unidos o en Europa.

Uno de los objetivos de nuestro proyecto era tratar de aportar más datos sobre esta cuestión y, especialmente, poder hacer un dibujo más cercano a los editores de la Wikipedia española. Por eso, diseñamos un breve cuestionario, con preguntas de tipo sociodemográfico pero también sobre percepción de calidad de

¹⁹ *Wikipedia Editors Study. Results from the Editor Survey*, April 2011. Disponible en: https://upload.wikimedia.org/wikipedia/commons/7/76/Editor_Survey_Report_-_April_2011.pdf

Wikipedia, y lo distribuimos online entre una selección de editores. Los encuestados se seleccionaron a partir del registro de ediciones de la Wikipedia española. En concreto, se seleccionaron aquellos editores con más de 5.000 ediciones. Los datos recogidos con el cuestionario se han tratado de manera anónima.

El cuestionario se envió por correo electrónico directamente a estos editores, y finalmente se recibieron un total de 23 respuestas. El periodo de la encuesta fue de aproximadamente dos meses. En la siguiente tabla (Tabla 1), se muestra la correspondiente ficha técnica.

Tabla 28. Ficha técnica del cuestionario

<p>Ficha técnica:</p> <p>Selección de la muestra: cuestionario enviado a los editores más activos (> 5000 ediciones)</p> <p>Número de respuestas: 23</p> <p>Encuesta: https://docs.google.com/a/uoc.edu/forms/d/1ol0TS6CIsD2o9sJfwP7ArqDqDkBy2KzFnTYp5mjtCMs/edit?usp=drive_web)</p> <p>Tipología de la encuesta: on-line</p> <p>Periodo: respuestas recibidas entre el 30/11/2015 y el 19/01/2016</p>
--

Fuente: Elaboración propia

Características de los editores

En este apartado se describen las principales características sociodemográficas de los editores que han contestado la encuesta. En primer lugar, analizamos el género de los encuestados. Como podemos observar en la Tabla 2, Todos los editores que han contestado la encuesta son hombres. Este resultado está en consonancia con la situación general de Wikipedia, donde la gran mayoría de editores son masculinos. Sólo un 10% son mujeres, y desde la propia Wikimedia España se han llevado a cabo diferentes iniciativas para aumentar este porcentaje.

Tabla 29. Distribución de los encuestados por sexo

	<i>Frecuencia</i>	<i>Porcentaje</i>	<i>Porcentaje acumulado</i>
Hombre	23	100,0	100,0
Mujer	0	0	0

Fuente: Elaboración propia

Por lo que respecta al país de residencia (ver Tabla 3), la mayoría de los editores residen en España (43,4%). En segundo lugar, se encuentran los que residen en Argentina (21,7%). Excepto en dos casos (que residen en Alemania y Estados Unidos), todos son residentes en países de habla hispana.

Tabla 30. Distribución de los encuestados por país de residencia

	<i>Frecuencia</i>	<i>Porcentaje</i>	<i>Porcentaje acumulado</i>
Alemania	1	4,3	4,3
Argentina	5	21,7	26,1
Chile	1	4,3	30,4
España	9	39,1	69,6
España - Venezuela	1	4,3	73,9
Estados Unidos	1	4,3	78,3
México	3	13,0	91,3
Uruguay	1	4,3	95,7
Venezuela	1	4,3	100,0
Total	23	100,0	

Fuente: Elaboración propia

En cuanto a su formación (ver Tabla 4), los editores encuestados tienen mayoritariamente una formación universitaria, de segundo y tercer ciclo (73,9%). En el caso de los editores con formación universitaria, hay una diversidad muy grande de especialidades (ver Tabla 5), aunque la mayoría son de Ciencias o Ciencias Sociales.

Tabla 31. Distribución de los encuestados por formación

	<i>Frecuencia</i>	<i>Porcentaje</i>	<i>Porcentaje acumulado</i>
Doctorado	4	17,4	17,4
Educación secundaria	1	4,3	21,7
Formación profesional no universitaria	5	21,7	43,5
Grado/licenciatura universitaria	8	34,8	78,3
Máster universitario	5	21,7	100,0
Total	23	100,0	

Fuente: Elaboración propia

Tabla 32. Distribución de los encuestados por especialidad (formación)

	<i>Frecuencia</i>	<i>Porcentaje</i>	<i>Porcentaje acumulado</i>
Ns/Nc	7	30,4	30,4
Administración	1	4,3	34,8
Arquitectura	1	4,3	39,1
Ciencias de la Computación	1	4,3	43,5
Ciencias Geológicas (Paleontología)	1	4,3	47,8
Derecho	1	4,3	52,2
Enfermería	1	4,3	56,5
Estadística-matemática	1	4,3	60,9
Filosofía	1	4,3	65,2
Física	2	8,7	73,9
Geography and Social Sciences	1	4,3	78,3
Ingeniería agronómica, Física de suelos, Agroclimatolo	1	4,3	82,6
Música	1	4,3	87,0
Profesorado de Inglés	1	4,3	91,3
Psicología	1	4,3	95,7
Química	1	4,3	100,0
Total	23	100,0	

Fuente: Elaboración propia

Finalmente, respecto la ocupación de los editores encuestados, también se observa una gran diversidad de resultados (ver Tabla 6). De todas formas se puede observar que hay un 21,5% que están jubilados/retirados.

Tabla 33. Distribución de los encuestados por ocupación

	<i>Frecuencia</i>	<i>Porcentaje</i>	<i>Porcentaje acumulado</i>
Ns/Nc	3	13,0	13,0
Actualmente sin trabajo	1	4,3	17,4
Asesoría, pero actualmente jubilado	1	4,3	21,7
Computista	1	4,3	26,1
Consulta privada	1	4,3	30,4
Docente	1	4,3	34,8
<i>Emeritus professor</i>	1	4,3	39,1
Empleado	1	4,3	43,5
Empleado público	1	4,3	47,8

Enfermero de Cuidados Intensivos	1	4,3	52,2
Informático teórico y artista escénico	1	4,3	56,5
Investigador	1	4,3	60,9
Jubilado (fui profesor de la ETSAM y arquitecto)	1	4,3	65,2
Maestro (jubilado)	1	4,3	69,6
No	1	4,3	73,9
Preparando proyecto.	1	4,3	78,3
Profesor de inglés y traductor	1	4,3	82,6
Profesor de Secundaria	1	4,3	87,0
Programador	1	4,3	91,3
retirado	1	4,3	95,7
Traductor y músico	1	4,3	100,0
Total	23	100,0	

Fuente: Elaboración propia

Análisis de la percepción de la calidad de Wikipedia

Para analizar la percepción de calidad de los editores más frecuentes de Wikipedia, hemos preguntado sobre tres aspectos fundamentales: su fiabilidad, su completitud y su actualización. Sobre una escala de Likert de 5 puntos (“1” = poco fiables/muy incompletos/poco actualizados; “5” = muy fiables/muy completos/muy actualizados), se plantearon las siguientes preguntas:

- En general, ¿crees que los artículos de Wikipedia son fiables?
- En general, ¿crees que los artículos de Wikipedia son completos?
- En general, ¿crees que los artículos de Wikipedia están actualizados?

A partir de los resultados que se muestran en la Tabla 7, se puede observar que los editores encuestados consideran que la calidad de la edición española de Wikipedia es bastante buena. Los tres indicadores de calidad (fiabilidad, completitud y actualización) están por encima de 3, el punto medio en el rango de posibles respuestas. De los tres aspectos, el mejor valorado es la actualización. El peor valorado es la completitud, dónde además se observa que ningún editor ha dado ni la mínima puntuación, “1”, ni la máxima “5”.

Tabla 34. Análisis descriptivo de la percepción de calidad

	<i>N</i>	<i>Mínimo</i>	<i>Máximo</i>	<i>Media</i>	<i>Desviación estándar</i>
Fiable	23	1,00	4,00	3,5652	,78775
Completa	23	2,00	4,00	3,1739	,71682
Actualizada	23	1,00	5,00	3,6957	,82212
N válido (por lista)	23				

Fuente: Elaboración propia

Seguidamente vamos a analizar la percepción de calidad según la formación de los editores y según su ocupación. Como en el caso general, se observa que el nivel de calidad percibido es bastante bueno, ya que todos los valores para los diferentes indicadores de calidad están por encima del punto medio, “3” (ver Tabla 8). Para aquellos editores que tienen una formación de doctorado o de postgrado, los indicadores de calidad que tienen una valoración más elevada son la fiabilidad y la actualización. Para el resto de casos, el indicador más valorado es solamente la actualización. El aspecto peor valorado continúa siendo, igual que en el caso general, la completitud.

Tabla 35. Percepción de la calidad según la formación de los editores

		<i>N</i>	<i>Media</i>	<i>Desviación estándar</i>	<i>Error estándar</i>
Fiable	Doctorado/Postgrado	7	3,7143	,48795	,18443
	Grado	10	3,4000	1,07497	,33993
	Otros	6	3,6667	,51640	,21082
	Total	23	3,5652	,78775	,16426
Completa	Doctorado/Postgrado	7	3,1429	,69007	,26082
	Grado	10	3,1000	,73786	,23333
	Otros	6	3,3333	,81650	,33333
	Total	23	3,1739	,71682	,14947
Actualizada	Doctorado/Postgrado	7	3,7143	,48795	,18443
	Grado	10	3,5000	1,08012	,34157
	Otros	6	4,0000	,63246	,25820
	Total	23	3,6957	,82212	,17142

Fuente: Elaboración propia

De todas formas, no podemos afirmar que haya diferencias estadísticamente significativas entre los tres colectivos analizados. Según el resultado del análisis ANOVA realizado (ver Tabla 9), se observa que los valores-*p* asociados a cada indicador de calidad están claramente por encima de 0,05.

Tabla 36. Análisis ANOVA para la diferencia de la percepción de calidad según la formación

		<i>Suma de cuadrados</i>	<i>gl</i>	<i>Media cuadrática</i>	<i>F</i>	<i>Sig.</i>
Fiable	Entre grupos	,490	2	,245	,372	,694
	Dentro de grupos	13,162	20	,658		
	Total	13,652	22			
Completa	Entre grupos	,214	2	,107	,193	,826
	Dentro de grupos	11,090	20	,555		
	Total	11,304	22			
Actualizada	Entre grupos	,941	2	,470	,676	,520
	Dentro de grupos	13,929	20	,696		
	Total	14,870	22			

Fuente: Elaboración propia

También, como en el caso general, se observa que el nivel de calidad percibido para los jubilados/parados y los ocupados es bastante bueno, ya que todos los valores para los diferentes indicadores de calidad son iguales o están por encima del punto medio, “3” (ver Tabla 10). Los editores ocupados son los que dan una valoración más elevada en los tres indicadores de calidad. La completitud continúa siendo el indicador peor valorado por parte de los dos colectivos, mientras que la actualización de los artículos también continúa siendo el aspecto mejor valorado por ambos grupos.

Tabla 37. Percepción de la calidad según la ocupación de los editores

		<i>N</i>	<i>Media</i>	<i>Desviación estándar</i>	<i>Error estándar</i>
Fiable	jubilados/parados	8	3,2500	1,16496	,41188
	ocupados	12	3,7500	,45227	,13056
	Total	20	3,5500	,82558	,18460
Completa	jubilados/parados	8	3,0000	,75593	,26726
	ocupados	12	3,4167	,66856	,19300
	Total	20	3,2500	,71635	,16018
Actualizada	jubilados/parados	8	3,5000	1,19523	,42258
	ocupados	12	3,9167	,51493	,14865
	Total	20	3,7500	,85070	,19022

Fuente: Elaboración propia

De todas formas, como en el caso de la formación, no podemos afirmar que haya diferencias estadísticamente significativas entre los dos colectivos analizados. Según el resultado del análisis ANOVA realizado (ver Tabla 11), se observa que los valores-p asociados a cada indicador de calidad están claramente por encima de 0,05.

Tabla 38. Análisis ANOVA para la diferencia de la percepción de calidad según la ocupación

		<i>Suma de cuadrados</i>	<i>gl</i>	<i>Media cuadrática</i>	<i>F</i>	<i>Sig.</i>
Fiable	Entre grupos	1,200	1	1,200	1,838	,192
	Dentro de grupos	11,750	18	,653		
	Total	12,950	19			
Completa	Entre grupos	,833	1	,833	1,682	,211
	Dentro de grupos	8,917	18	,495		
	Total	9,750	19			
Actualizada	Entre grupos	,833	1	,833	1,161	,295
	Dentro de grupos	12,917	18	,718		
	Total	13,750	19			

Fuente: Elaboración propia

Recursos y otras ediciones de Wikipedia

Además de la edición en lengua española, la mayoría de los editores también editan en otras ediciones de Wikipedia (ver Tabla 12). Gran parte de ellos, un 47,83%, también editan en la versión en inglés. Un 30,43% reconoce que sólo edita en la versión española.

Tabla 39. Otras ediciones de Wikipedia

	<i>Frecuencia</i>	<i>Porcentaje</i>	<i>Porcentaje acumulado</i>
Ns/Nc	3	13,0	13,0
de, en, fr	1	4,3	17,4
En pocas ocasiones.	1	4,3	21,7
English	1	4,3	26,1
English - Catalan	1	4,3	30,4
Inglés	1	4,3	34,8
Inglés muy esporádicamente	1	4,3	39,1
inglés, italiano, catalán, portugués	1	4,3	43,5
Inglés, italiano, portugués	1	4,3	47,8
Inglés, portugués y francés	1	4,3	52,2
ninguna otra	1	4,3	56,5
no	2	8,7	65,2
No	5	21,7	87,0
Sobre todo en la Wikipedia en inglés, pero también ediciones puntuales en otras como las Wikipedias en francés, en alemán y en portugués, entre o	1	4,3	91,3

Wikipedia en francés, inglés, portugués, pero mucho menos que en español	1	4,3	95,7
Wikipedia en inglés y Wikimedia Commons	1	4,3	100,0
Total	23	100,0	

Fuente: Elaboración propia

Respecto las fuentes de información que utilizan los editores para realizar sus aportaciones, no hay un consenso entre las respuestas realizadas. Como puede comprobarse en la Tabla 13, hay una gran diversidad de recursos y de fuentes de información: artículos científicos, otras ediciones de Wikipedia (principalmente la versión en inglés), libros, Internet, etc. La principal conclusión que se puede extraer de estos resultados es que prácticamente todos los editores tienen en cuenta otras fuentes de información para documentarse adecuadamente a la hora de hacer sus contribuciones.

Tabla 40. Fuentes Información

	<i>Frecuencia</i>	<i>Porcentaje</i>	<i>Porcentaje acumulado</i>
Ns/Nc	2	8,7	8,7
Artículos arbitrados, libros impresos o electrónicos.	1	4,3	13,0
Artículos científicos	1	4,3	17,4
conocimiento propio, libros, internet	1	4,3	21,7
Google books	1	4,3	26,1
Google y Wikipedia en otros idiomas	1	4,3	30,4
gráficos y wikipedia inglesa	1	4,3	34,8
Internet, revistas científicas	1	4,3	39,1
IUCN, SEO/Birdlife, Avibase etc	1	4,3	43,5
Libros de Google books	1	4,3	47,8
Libros en versión digital o ninguna, pues mi trabajo no siempre es con el contenido sino también con categorización, imágenes y sobre todo có	1	4,3	52,2
Libros, publicaciones, e información fiable en línea	1	4,3	56,5
Libros, sitios web formales y oficiales	1	4,3	60,9
Mi saber y lo que pueda necesitar de mi biblioteca o de internet.	1	4,3	65,2
Otras wikipedias y trabajo de recabar referencias	1	4,3	69,6
Periódicos y sitios web (no blogs ni foros)	1	4,3	73,9
Publicaciones científicas, filosóficas, prensa, etc.	1	4,3	78,3
Revisats científicas en la Web	1	4,3	82,6
Revistas científicas, prensa especializada, etc.	1	4,3	87,0
Traduzco	1	4,3	91,3
Variadas	1	4,3	95,7

Wikipedia en inglés, libros y documentación técnica, Internet	1	4,3	100,0
Total	23	100,0	

Fuente: Elaboración propia

Dada la importancia de los recursos y las fuentes de información a la hora de editar en Wikipedia, se preguntó a los editores sobre si consideraban que Wikipedia debería proporcionar más recursos o información para poder editar artículos de ciencia y tecnología. Un 34,8% de los editores considera que no es necesario que Wikipedia aporte más recursos para la edición, mientras que un 43,5% se muestra favorable a esta iniciativa. El resto, NS/NC.

5.2. Estudio del proceso de edición

El estudio del proceso de edición de los artículos científicos se ha llevado a cabo mediante la comparación entre distintas métricas para los artículos del Corpus y para el conjunto de la Wikipedia española. Concretamente, se han comparado el tamaño de los artículos (en bytes), el número de editores y el número de ediciones.

La base de datos proviene de un archivo generado, a partir del mismo *dump* de Wikipedia descargado inicialmente (en diciembre de 2014), con todas las páginas de Wikipedia (no solo artículos, sino historiales, etc.). Este archivo recogía, para cada revisión realizada en un artículo, información acerca de (1) la página donde se había hecho la revisión, (2) la fecha en que tuvo lugar, (3) el usuario que la realizó, (4) los bytes que tras la edición se añadían o se restaban a la versión anterior del artículo, (5) si es o no una revisión menor –si bien este dato no es del todo fiable, porque lo pone el mismo usuario–, (6) si es una revisión a una página del Corpus, y (7) si la página es del Corpus, a qué comunidad (según la Clasificación UNESCO a dos dígitos) pertenece.

Hay que atender unas consideraciones técnicas previas. En primer lugar, la mayoría de ediciones provienen de usuarios anónimos (de los que queda en el registro una dirección IP) y no de usuarios registrados. Dado que un usuario puede tener distintas direcciones IP, se puede haber introducido un cierto sesgo al considerar que cada dirección IP identifica a un editor. En segundo lugar, se han eliminado las ediciones realizadas por *bots*, habiendo identificado estos *bots* a partir de una lista oficial (con un total de 432), pero puede haber *bots* que no consten en esta lista oficial y que, por lo tanto, no han sido eliminados del archivo. También se han eliminado páginas con contenidos concretos (p.e. páginas de usuarios, discusiones, páginas de pruebas, etc.).

El análisis se centra en cuatro puntos:

a) Comparación global entre la Wikipedia española y el Corpus: Los resultados indican que en los artículos del Corpus hay más ediciones, hechas por más personas, y más largas que en la Wikipedia española en general. Ponderando por el número de artículos, observamos también que las páginas del Corpus tienen más editores por página.

En números absolutos, Wikipedia tiene un total de 6.801.427 editores, de los cuales el 8'9% son usuarios registrados (6195967 usuarios anónimos –IPs– y 605460 usuarios registrados). El Corpus cuenta con 1.003.250 editores, de los cuales el 9'5% son usuarios registrados (908.383 usuarios anónimos –IPs– y 94.867 usuarios registrados). A continuación se muestra una tabla resumen de los resultados en números absolutos.

Tabla 41. Editores, ediciones y bytes editados, en es.W y en el Corpus. Comparativa.

	Wikipedia		Corpus	
Editores	Usuarios anónimos	6.195.967 IPs	Usuarios anónimos	908.383 Ips
	Usuarios registrados	605.460 (8'9%)	Usuarios registrados	94.867 (9'5%)
	Total	6.801.427	Total	1.003.250

Ediciones	49.728.066		3.882.551	
Bytes editados	Ediciones positivas	31.577.053 (1352595993 bytes)	2299317 (1096611368 bytes)	
	Ediciones negativas	14363618 (7098183205 bytes)	1353339 (839095656 bytes)	
	Ediciones neutras	3.787.395	229.895	
“Densidad” de editores	6801427 / 1688187 = 4’03 editores por página		1003250 / 59870 = 16’76 editores por página	

Fuente: Elaboración propia

b) Distribución de editores en función del tipo de artículos que editan: El objetivo de este análisis es distinguir entre los editores de Wikipedia que también editan en el Corpus (‘mixtos’), editores que no editan el Corpus (‘acientíficos’) y aquellos que solo editan el Corpus (‘científicos’). Los resultados nos indican que los editores más activos del Corpus lo son también en la Wikipedia en general. Curiosamente, aquellos que editan solo el Corpus son los menos activos.

Todos los editores registrados en el Corpus están, obviamente, en Wikipedia. Los editores registrados del Corpus son el 15’67% de los editores registrados en Wikipedia, y hacen el 84’40% (25.872.214) de todas las ediciones hechas por el total de usuarios registrados a Wikipedia.

En Wikipedia, el 7’24% de los editores registrados hace el 92’76% de todas las ediciones. Redondeando, poco más del 7% de los editores hace 29 ediciones o más. En el Corpus, el 9’6% de los editores registrados hace el 90’4% de todas las ediciones. Redondeando, el 9’7% de los editores hace 13 ediciones o más. Por lo tanto, en el Corpus hay más “participación” (no está tan concentrada).

Por otro lado, el Corpus lo editan 94.867 usuarios registrados, con un total de 2.210.943 ediciones en el Corpus (272’72 ediciones por editor). De estos, 39.764 (un 41’9%) solo editan el Corpus, pero solo con 97.129 ediciones (solo un 4’4% de todas las ediciones. Es decir, los editores exclusivos del Corpus son poco activos, con 2’4 ediciones por editor). Al contrario, los editores del Corpus que también editan la Wikipedia (55.103, un 58’1% de los editores del Corpus), lo hacen 23.661.271 veces (429’40 ediciones por editor en la Wikipedia), mientras que los 580.691 editores registrados en Wikipedia que no editan el Corpus, solo hacen 5.150.750 ediciones (8’87 ediciones por editor).

Tabla 42. Ediciones y tipos de editores, en es.W y en el Corpus

	<i>Editors</i>	<i>Ediciones en el Corpus</i>	<i>Ediciones en Wikipedia</i>
Editores científicos	39.764	97.129	0
Editores mixtos	55.103	2.113.814	23.661.271
Editores acientíficos	580.691	0	5.150.750

Fuente: Elaboración propia

c) Comparación de las páginas editadas con el número de editores, en Wikipedia y en el Corpus (totales y distinguiendo entre anónimos y registrados). Según los resultados, los usuarios registrados son mucho más prolíficos que los no registrados, tanto en el Corpus como en Wikipedia.

Tabla 43. Ediciones y tipos de usuarios, en es.W y en el Corpus

	Wikipedia	Corpus
Usuarios anónimos	18.705.102 3'02 ediciones por IP	1.671.608 1'84 ediciones por IP
Usuarios registrados	31.022.964 51'24 ediciones por usuario registrado	2.210.943 23'31 ediciones por usuario registrado

Fuente: Elaboración propia

d) Comparación del volumen de edición (en bytes), entre usuarios anónimos y registrados, tanto en Wikipedia como en el Corpus: Los usuarios registrados hacen ediciones más largas, tanto en el Corpus como en Wikipedia. Las ediciones en Wikipedia son casi el doble de largas (en bytes) que en el Corpus, en lo que respecta a los usuarios registrados.

En Wikipedia, los usuarios anónimos editan un total (sumando ediciones positivas y negativas) de 6.687.482.878 bytes. Esto supone 1079'33 bytes por usuario anónimo. En cambio, los usuarios registrados editan un total de 13.936.660.320 bytes (23.018'30 bytes por usuario registrado).

En el Corpus, los usuarios anónimos editan un total (sumando ediciones positivas y negativas) de 799.182.952 bytes (879'79 bytes por usuario anónimo), mientras que los usuarios registrados editan un total de 1.136.524.072 bytes (11.980'18 bytes por usuario registrado).

6. Ciencia y contexto social

A pesar del deliberado carácter lego de sus contribuciones y de la estrategia colaborativa y abierta de la participación en Wikipedia, existen fuertes paralelismos entre su proceso de edición (similar, por cierto, al de otras iniciativas de producción entre iguales basadas en el procomún) y el proceso habitual de elaboración del conocimiento científico.

En primer lugar, Wikipedia descansa, por encima de todo, en un sistema de revisión por pares muy estricto: el destino de cualquier aportación (edición) depende de lo que hagan y piensen los demás usuarios - algo muy similar a lo que sucede en la ciencia. La gran diferencia es que los pares son, en principio, el resto de miembros de la comunidad de lectores y editores o, mejor dicho, cualquier persona que quiera llevar a cabo esta tarea y no un conjunto predeterminado de especialistas acreditados como tales.

El uso de referencias bibliográficas es también similar. Ya hemos mencionado que una de las reglas básicas en la edición de un artículo en Wikipedia es que toda afirmación debe estar convenientemente apoyada una fuente fiable y externa a la propia enciclopedia. El uso sistemático de referencias para indicar las fuentes de trabajos anteriores es, también, uno de los atributos básicos de la literatura científica.

A pesar de la ausencia de una estructura jerárquica formal de mando o de gobernanza en Wikipedia, los miembros de la comunidad pueden aspirar a ocupar determinados roles de responsabilidad que los habilitan para tareas de mantenimiento o de lucha contra el vandalismo - en particular, pueden llegar a convertirse en “burócratas” o “bibliotecarios”, en la versión española. Estas categorías, aunque no implican ninguna autoridad per se, se consiguen meritocráticamente a partir de una cierta experiencia previa en la edición de artículos y de la confianza otorgada por otros miembros de la comunidad. Como en el terreno de la ciencia, la progresión depende básicamente de los méritos pasados.

Pero a pesar de estas similitudes tan claras, existen también diferencias notables entre el *modus operandi* característico de la mayoría de ciencias y el de Wikipedia. En especial, el objetivo de Wikipedia, como enciclopedia, es recoger, de forma sintética y accesible para el público general, conocimiento consolidado, es decir, aceptado por la comunidad científica - si se trata de temas científicos - y, por tanto, previamente publicado en fuentes primarias. Este aspecto esencial queda recogido en una norma oficial básica de Wikipedia conocida en inglés como “*no original research*”.

Encontramos otras diferencias notorias en el proceso de revisión por pares (que en la ciencia es *previo* a la publicación mientras que en Wikipedia es *posterior* a la publicación), en la capacidad de participación en este proceso (ya que en el ámbito científico hay que ser miembro de la comunidad científica y reunir ciertas características para poder ser revisor, mientras que en Wikipedia es accesible a cualquier usuario) y en el proceso de autoría de los artículos (que en Wikipedia es colectiva).

En realidad, en ciencia lo único que acaba siendo verdaderamente público es el resultado final: el artículo modificado por el autor de acuerdo con los requerimientos de los revisores y editores. En cambio, el proceso de

revisión en Wikipedia permite establecer un registro de los editores/revisores (aunque eso no signifique poder conocer su “identidad”), así como de las modificaciones, los comentarios e incluso el historial de ediciones.

Por otro lado, en Wikipedia no es necesario ningún tipo acreditación formal para participar y contribuir. De hecho, Wikipedia se presenta a menudo como "la enciclopedia libre que todos pueden editar". A la hora de juzgar una contribución concreta, las credenciales formales del editor son irrelevantes, por lo que el crédito o reputación en Wikipedia se construye únicamente por lo que se hace dentro de Wikipedia. La ausencia de este tipo de acreditaciones no significa que la comunidad colabore de forma anárquica. Wikipedia posee, en realidad, una organización muy sofisticada que ha ido construyendo un complejo entramado de reglas y normas, frecuentemente citadas por los editores en las disputas sobre calidad y la resolución de conflictos.

6.1. Estudio de las páginas de discusión

A pesar de que Wikipedia es a menudo presentada como un ejemplo de “democratización del conocimiento” – y, en cierto sentido, puede entenderse así - una de sus políticas menos conocidas reza “Wikipedia no es una democracia”. En efecto, la inmensa mayoría de procesos implicados en la elaboración de Wikipedia no se basan en procedimientos democráticos. Ni siquiera la resolución de conflictos de edición. Estos conflictos, aunque poco habituales considerando el volumen de ediciones que se producen constantemente, implican discrepancias sobre lo que un artículo concreto debería decir o no decir (sobre la fiabilidad de un dato o afirmación, por ejemplo).

En el ámbito de la ciencia, las controversias son también habituales, y tampoco se resuelven mediante procedimientos democráticos sino por un sistema similar de aportación de nuevas evidencias o argumentos. Como en el caso de Wikipedia, no siempre las controversias terminan en el consenso entre las partes implicadas (Collins & Pinch, 1993) y pueden perdurar durante largos períodos de tiempo sin clausurarse.

Las discusiones (en el sentido más amplio del término) son un elemento característico de Wikipedia, en tanto que el proceso de creación de los artículos es constante y muy dinámico, del mismo modo que las actividades de la comunidad y la colaboración entre sus miembros se extienden más allá de la creación de artículos enciclopédicos. De hecho, todos los artículos cuentan con un espacio específico donde poder debatir el desarrollo de los contenidos de los textos: las páginas de discusión. Estos espacios se usan de manera muy diversa y, por lo tanto, dan lugar a discusiones muy variadas.

Así como desde los estudios de ciencia y tecnología (STS) se considera que el estudio de las controversias científicas permite observar procesos clave en el desarrollo de las ciencias, nuestro proyecto supone que a partir del estudio de las páginas de discusión de artículos de ciencia y tecnología se podrá observar hasta qué punto en Wikipedia tales procesos difieren del entorno académico o hasta qué punto se reproducen los mismos esquemas de intercambio y crítica.

La muestra de páginas de discusión a analizar se creó a partir del Corpus del proyecto, puesto que de las 60.108 páginas del Corpus, 19.492 eran páginas de discusión. Para cada página de discusión se almacenó la información acerca del número de editores que participaron en ella, el total de ediciones, la fecha de creación de la página, la fecha de la última edición y el tamaño de la página (en bytes).

Para determinar la relevancia de las páginas de discusión para el análisis, decidimos no tener en cuenta únicamente el tamaño de la discusión (a partir de la literatura existente y los resultados de una primera aproximación a la lista de páginas ordenadas de mayor a menor tamaño), sino también el número de editores y ediciones. Mirando la distribución de estas tres variables, decidimos descartar las páginas con un tamaño inferior a 1K, con solo un editor y con una sola edición. Aplicando este criterio el conjunto de datos se redujo a 4.634 páginas de discusión.

Para ordenar este conjunto de páginas de discusión por relevancia, a partir de la combinación de las tres variables mencionadas (tamaño, editores y ediciones), se realizó un análisis de componentes principales (PCA). Con las tres variables normalizadas, decidimos usar la ordenación del componente 1 porque explicaba el 77% de los casos. En un extremo situaba las páginas con pocos editores, pocas ediciones y menor tamaño. En el otro extremo, lo contrario: muchos editores, muchas ediciones y mayor tamaño.

Con una selección de 25 páginas de discusión, de entre las primeras páginas de la lista creada con el PCA, se llevó a cabo un análisis cualitativo con un modelo de análisis propio. Los resultados nos permiten inferir algunos comportamientos interesantes.

En muchas de las páginas de discusión no se da realmente una discusión científica (no se discute sobre aspectos como la veracidad o fiabilidad de los contenidos del artículo), sino que se debaten aspectos técnicos como corrección de errores, sugerencias de mejora de los apartados del artículo, o dudas acerca de cuestiones muy concretas que se quieren resolver. No hay, por lo tanto, controversias científicas en la mayoría de páginas de discusión que hemos estudiado.

Sin embargo, cuando la discusión aparece, los temas son recurrentes:

a) **Neutralidad:** La falta de neutralidad es una de las demandas que más aparece en caso de discusión. Cuando esto ocurre, un editor considera que una de las informaciones contenidas en el texto no es neutral, es decir, responde a una opinión particular de la persona que lo ha editado. A partir de aquí, la situación puede tomar distintas direcciones, según a qué se refiera la falta de neutralidad:

- Puede ser que se acuse al otro de no haber interpretado bien las fuentes originales y se saque a relucir la necesidad de tener unos conocimientos determinados para poder elaborar correctamente un artículo científico. En estos casos, la falta de neutralidad se atribuye a problemas de comprensión de las fuentes originales, del mismo asunto del que se escribe, o de las cuestiones subyacentes en los temas controvertidos. La falta de neutralidad es aquí un problema de interpretación. Por ejemplo, en el artículo ‘Método científico’ se discute sobre la “traducción libre” de las obras de Popper y otros autores anglosajones (falsación, probatoriedad...). O también, en el artículo ‘Efecto Coriolis’, se comenta “¿Por qué los que no tienen una formación en física hacen afirmaciones tan tajantes y falsas?”.

- Mucho más habitual, en cambio, es relacionar la falta de neutralidad con una posición parcial. Es decir, se considera que la información editada no es neutral porque responde a una opinión personal. Se duda de la neutralidad de las aportaciones cuando contradicen las opiniones propias. Entonces, los partidarios y detractores de una “idea” se acusan mutuamente de falta de neutralidad. Por ejemplo, en el artículo ‘Sigmund Freud’ aparece “se han eliminado opiniones personales a favor del psicoanálisis quizás demasiado parciales, pero quedando todo al final en un antiparcialismo extremo”. O también “Se trata de una especie de reflexión pseudo-metafísica que no sé muy bien a cuento de qué viene, que es personal de su autor y no aporta información alguna al artículo” (Página de discusión del artículo ‘Homo Sapiens’).

En estos casos, el debate sobre la neutralidad se relaciona a menudo con un debate sobre la evidencia científica disponible para hacer aseveraciones sobre la realidad. Aquí la discusión se construye sobre el eje opinión/hecho, siendo la opinión “construida” y un hecho “la verdad”; siempre se puede separar el hecho científico de posiciones políticas, ideológicas o personales. Por ejemplo, en el artículo ‘Alimento orgánico’, un editor responde “lo que yo veo que va a pasar a la larga es que tus interpretaciones no van a quedar, solo vas a conseguir que aparezcan por un tiempo: días o meses quizás, pero la ciencia nos dá la razón y más temprano que tarde te van a demostrar que estás introduciendo parcialismos “interesados”, lo sabes perfectamente.”

- Más aún, se debate también la ‘captura’ de los científicos por parte de intereses espurios. Aquí el problema recae en que no siempre se considera que los hechos científicos sean una muestra probatoria de la realidad. Por ejemplo, en la página de discusión del artículo Efecto invernadero aparece el siguiente comentario: “Que la mayoría de los países (reunidos en la ONU) sigan esta teoría del calentamiento global (...) no quiere decir que la teoría sea cierta”. O también, en la página de discusión del artículo Alimento transgénico: “Los alimentos transgénicos fueron creados por corporaciones para así poder patentar algunas semillas (...) fin económico (...) porque ya está demostrado que producen tumores malignos y malformaciones en las personas”.

b) Necesidad de **referenciar** la información: Tanto en las discusiones sobre neutralidad como en otros casos, es muy habitual mencionar la necesidad de referenciar la información, añadiendo citas y bibliografía. Por ejemplo, “no está mal, pero deberías añadir alguna referencia, porque algún párrafo parece fuente primaria o investigación original” (Página de discusión del artículo ‘Método científico’). Cuanto más acalorada es una discusión, más referencias se piden y, al mismo tiempo, se cuestionan. Entonces, se llega a discutir sobre la veracidad de las fuentes aportadas, es decir, no solo es necesario aportar fuentes y referenciar la información, sino que no todas las fuentes de información son válidas. Por ejemplo, en ‘Psicología’ encontramos “Los diccionarios no son fuentes académicamente adecuadas para indicar que la Psicología es una ciencia (...) Este artículo fue escrito por Christian Nordqvist, ¿quién es el?, ¿es una autoridad en la materia? ¿quién me asegura eso? ¿es conocido en la comunidad científica? Lo digo porque su artículo no tiene fuentes que refuercen lo que dice”.

En este sentido, cabe destacar que las fuentes académicas suelen gozar de mayor prestigio que otras fuentes como publicaciones institucionales, revistas o prensa. A menudo, la demanda de nuevas fuentes es un recurso

para tratar de concluir una disputa, pero no consigue ese fin: cuando hay posiciones contrapuestas, se cuestiona también la fiabilidad de las fuentes que aportan los otros en el debate. De este modo, puede llegarse a una retahíla de informaciones contrapuestas que roza lo absurdo, como llegó a plasmar un editor que exasperó “¿Si digo que la Tierra es redonda, también me pediréis referencias?!” (Página de discusión del artículo ‘Alimento orgánico’).

c) Mención a las **normas de funcionamiento de Wikipedia**: La mención a normas de funcionamiento de Wikipedia es un recurso habitual en las discusiones. En especial, se hace referencia a las políticas de Punto de vista neutral, Verificabilidad y Fuentes fiables. Por ejemplo, “He añadido plantillas de "sin referencias" a una afirmación que efectivamente, sigue sin referencias” (Página de discusión del artículo ‘Alimento transgénico’). O también “Ustedes exponen que la etiqueta de *Este artículo o sección necesita referencias que aparezcan en una publicación acreditada, como revistas especializadas, monografías, prensa diaria o páginas de Internet fidedignas*. Ya les envié un dibujo explicativo (...)” (Página de discusión del artículo ‘Número áureo’).

Esta observación refuerza la tesis de Jemielniak (2014, p.105), según la cual en lugar de fomentar la confianza recíproca o asimétrica entre las personas (que en la mayoría de organizaciones suele apoyarse en credenciales formales), en Wikipedia se prefiere la confianza en los procedimientos y las normas.

Curiosamente, no hemos encontrado ninguna mención a la cuestión del contenido libre, el cuarto pilar. De manera indirecta, se debate también sobre qué tipo de debates deben tener lugar en los artículos de Wikipedia, haciendo así mención a otra norma básica: “Wikipedia es una enciclopedia”. Por ejemplo, “Este artículo ha caído en el problema de Wikipedia en inglés, en el cual la gente confunde un artículo enciclopédico con un fórum o un blog” (página de discusión del artículo ‘Efecto coriolis’). O “Esto no es un foro”, en la página de discusión del artículo ‘Teoría de cuerdas’.

Cuando la discusión es más acalorada, a menudo se producen descalificaciones entre los editores. Las posiciones contrarias tienden a concentrarse en dos editores enfrentados y aparecen algunos aliados que muestran su apoyo o disconformidad a los comentarios que se van sumando. Para resolver estos conflictos, la manera más habitual es recurrir a la intervención de bibliotecarios o administradores, usuarios con capacidad de bloqueo tanto de usuarios como de artículos. En cambio, en ningún caso se hacen votaciones como medida de resolución del conflicto. En realidad, la mayoría de discusiones observadas se terminan por agotamiento de alguna de las partes. Por ejemplo, “Yo me rindo. Hagan lo que crean con el artículo. Si piensan que solo es una apariencia del movimiento relativo pues lo ponen y ya está” (página de discusión del artículo ‘Efecto coriolis’). O “querido amigo, no hay nada más que aportar, ni este litigio ni tampoco en cuanto a referencias, pues están bastante completas. El problema radica en otro lugar” (Página de discusión del artículo ‘Sigmund Freud’).

En conclusión, hay indicios de que el comportamiento de los editores en las discusiones sobre el contenido de los artículos da lugar a cierta paradoja: por un lado, se refuerzan aspectos colaborativos como la libertad de participación, el trabajo colectivo, etc. Pero al mismo tiempo se pone mucho énfasis en aspectos más propios de la cultura científica tradicional (como la autoridad de la ciencia por encima de cualquier otro conocimiento) y con

una intensidad que no a menudo no se da en la misma academia. Los editores científicos de Wikipedia parecen, en este sentido, “más papistas que el Papa” si se nos permite la expresión.

6.2. Construcción de mapas cognitivos

Los mapas (cognitivos) de ciencias son representaciones simbólicas de los campos científicos u organizaciones en las que los elementos del mapa están asociados con los temas o áreas de conocimiento. Los elementos se colocan en el mapa de modo que los otros elementos con características relacionadas o similares se encuentran próximos, mientras que aquellos elementos que son distintos están colocados en lugares más distantes.

Los mapas se construyen sobre la base de una matriz de medidas de similitud, calculadas a partir de funciones de correlación entre los elementos de información presentes en diferentes elementos (por ejemplo, se puede considerar que dos artículos científicos estarán muy cerca si aparecen citados los mismos autores). Para una mejor interpretación de los resultados, las matrices multidimensionales se proyectan en dos o tres dimensiones.

La utilidad de los mapas radica en el hecho que permite explorar las relaciones entre diferentes elementos de manera sencilla. El análisis de estos mapas pertenece al campo del análisis de redes, y mejora la representación tradicional de la ciencia como un "árbol del conocimiento". Estos mapas permiten analizar la complejidad de las relaciones entre diferentes elementos, atendiendo a su interdisciplinariedad, multidisciplinariedad y transdisciplinariedad.

Aunque hay una larga tradición en la construcción de mapas mundiales de la ciencia, en la última década ha habido un crecimiento importante en este ámbito de conocimiento con la eclosión de la denominada “ciencia de los mapas”. Los mapas son muy útiles para la comprensión de la dinámica interna de un campo de investigación o disciplina emergente (potencialmente precisos en la descripción de las relaciones dentro de una misma área de conocimiento).

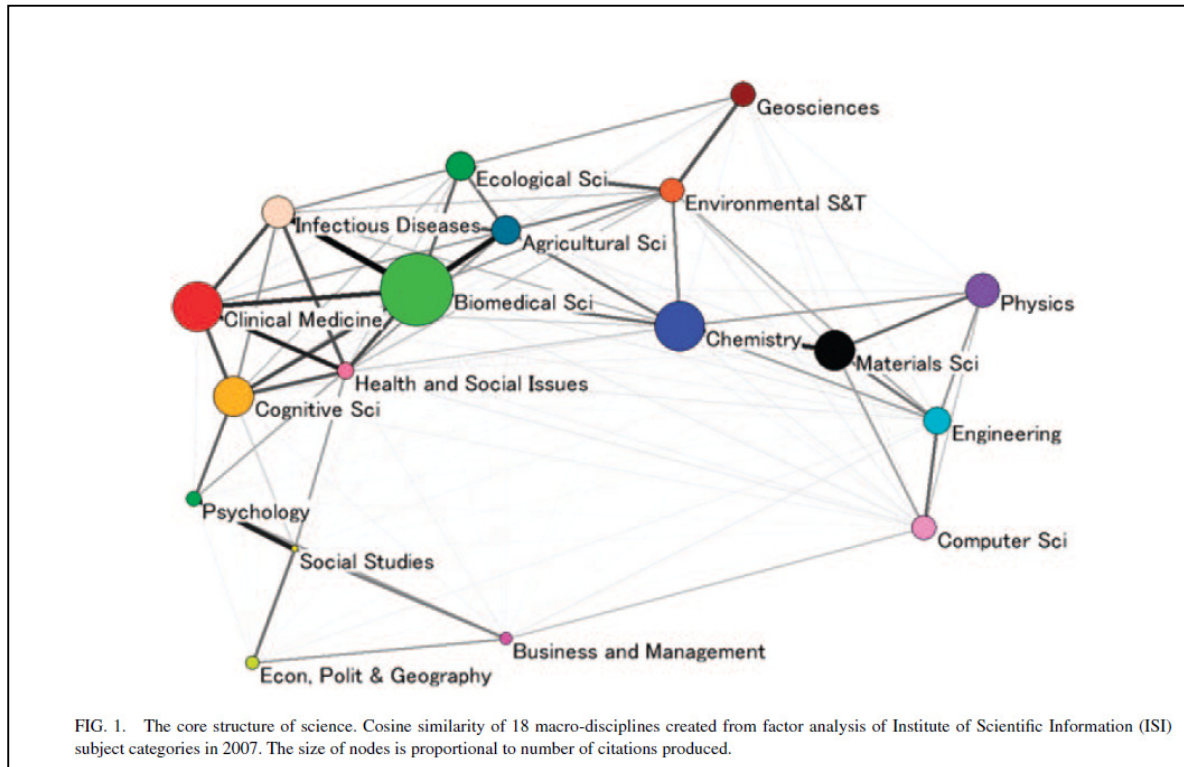
Diversos grupos de investigación independientes, usando diferentes opciones de análisis para medir la similitud entre ámbitos de conocimiento y considerando diversas técnicas de visualización, hay construido mapas de la ciencia con estructuras muy similares. Esto muestra que son representaciones sólidas de la ciencia.

Los principales elementos de consenso entre todas estas representaciones son las siguientes:

1. La ciencia no tiene una representación continua.
2. El mapa de la ciencia tiene una forma circular, donde ninguna disciplina es la dominante y ninguna ocupa el "centro". Cada disciplina puede ser considerada como el centro de ámbito de conocimiento.
3. Los efectos directos e indirectos entre disciplinas son similares.
4. Las distancias relativas entre las categorías deben ser interpretados con precaución, ya que la visualización de los mapas (en 2D o en 3D) implica necesariamente una simplificación de la realidad, con muchas más dimensiones.

La siguiente figura muestra uno de estos mapas de la ciencia, en la que puede apreciar la disposición circular de los diferentes ámbitos de conocimiento y la complejidad de las relaciones, directa e indirectas, entre disciplinas.

Figura 4. Mapa de la ciencia



Fuente: Elaboración propia

Mapa de la ciencia de la Wikipedia española

Teniendo en cuenta lo comentado en la introducción, en este apartado pretendemos realizar un mapa de la ciencia de la edición en español de Wikipedia y comprobar hasta qué punto coincide con las representaciones consensuadas a nivel global, como la de la anterior Figura 4. Siguiendo el procedimiento de construcción del Corpus de artículos de ciencia y tecnología de la Wikipedia española, los ámbitos reflejados en el mapa responden a la Clasificación de la Unesco, a dos dígitos.

Para calcular la similitud entre áreas de conocimiento, y de acuerdo con la literatura existente, podíamos considerar diferentes opciones. La gran mayoría de los mapas de la ciencia han utilizado alguna de las distintas unidades de análisis y medidas de similitud que se exponen a continuación:

1. Co-citaciones de artículos
2. Análisis *coword*
3. Co-clasificación de artículos
4. Co-citación de revistas científicas

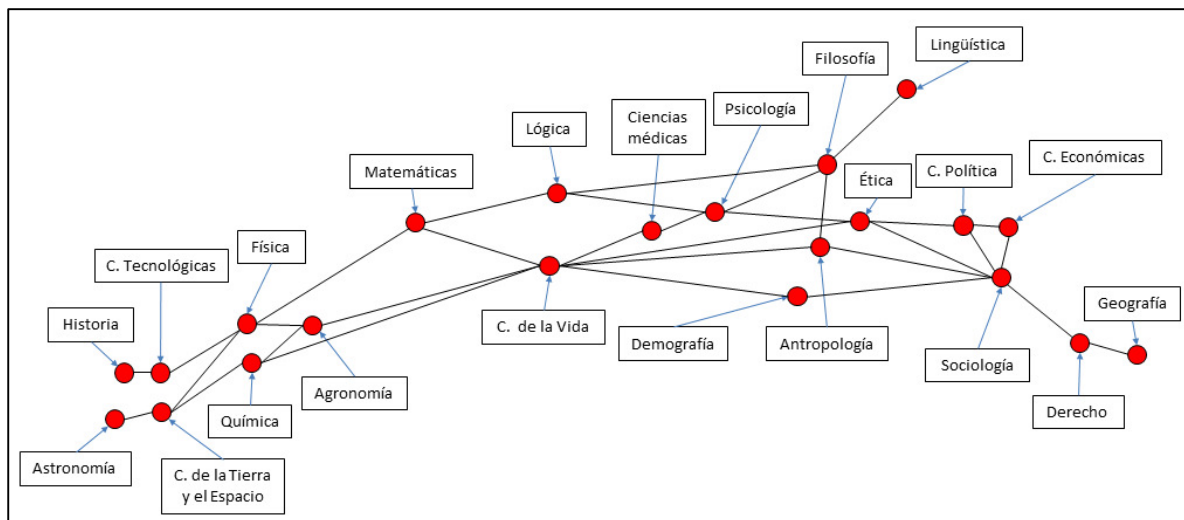
5. Co-citación de autores
6. ... (y combinaciones de las técnicas anteriores)

En nuestro caso hemos considerado una pequeña variación de las tres medidas de co-citación anteriores. En concreto hemos tomado como medida de similitud entre dos ámbitos de conocimiento la suma del número de enlaces que van de todos los artículos de Wikipedia de un ámbito de conocimiento a otro y viceversa. La cantidad de enlaces (en las dos direcciones) entre ámbitos nos da una idea de lo cercanos que pueden estar esos ámbitos de conocimiento. Si hay muy pocos artículos de un ámbito de conocimiento que enlacen a artículos de otro ámbito de conocimiento (y viceversa), tendremos una clara evidencia de que los ámbitos de conocimiento tienen pocos elementos en común, y por tanto en el mapa deberán situarse lejos el uno del otro.

Una vez hemos calculado el número de enlaces entre ámbitos, hemos construido la matriz de similitudes a partir del coseno que forman los vectores de enlaces de los ámbitos de conocimiento. Si el ángulo entre dos vectores es pequeño, esto nos indicará que los dos ámbitos de conocimiento están muy cerca (el coseno también será muy pequeño).

A partir de esta matriz de similitudes, y usando el programa PAJEK, hemos obtenido el siguiente mapa de la ciencia:

Figura 5. Mapa de la ciencia a partir del Corpus



Fuente: Elaboración propia

Si comparamos este resultado con el de la Figura 1, teniendo en cuenta que no tenemos exactamente la misma clasificación de ámbitos de conocimiento, podemos observar las siguientes importantes similitudes:

1. Las Ciencias Médicas y Ciencias de la Vida están en el "centro" de la representación.

Siguiendo el sentido de las agujas del reloj:

2. Psicología está muy cerca del ámbito de conocimiento de las Ciencias Médicas y del ámbito de conocimiento de las Ciencias de la Vida.

3. Seguidamente también se encuentra el grupo de los "estudios sociales": Demografía, Sociología y Antropología.
4. Y a continuación, Economía y Ciencia Política

A la izquierda del "centro" del mapa:

5. No muy lejos de las Ciencias Médicas y de la Vida, también nos encontramos con Física y de Química.
6. En la órbita de Física y Química, también encontramos un grupo similar de ámbitos de conocimiento: Agronomía, Ciencias de la Tierra y del Espacio, Astronomía y Astrofísica y Ciencias Tecnológicas.

7. Conclusiones

Este informe recoge los análisis y resultados más destacados de un estudio sobre el contenido científico de la Wikipedia, en su versión en español. Como tal constituye el primer estudio sistemático de dicho contenido, en cualquier versión lingüística de Wikipedia, realizado hasta el momento.

Una primera conclusión, de tipo metodológico, es la enorme dificultad de determinar con precisión el volumen de artículos de Wikipedia que tratan directamente temas de ciencia y tecnología. Gran parte de esta dificultad tienen que ver con la naturaleza particular de las “categorías” de Wikipedia y con sus características hipertextuales.

A pesar de ello, la técnica que hemos diseñado para determinar el “corpus científico” de Wikipedia, nos permite afirmar que alrededor de un 10% de los artículos en esta enciclopedia tratan directamente cuestiones científicas y tecnológicas – si bien creemos que una concepción más amplia de estas cuestiones (que incluyera obras artísticas, localizaciones geográficas – ciudades, países, etc. -, y otros elementos relacionados con determinados aspectos de las humanidades, las ciencias sociales y la tecnología) podría aumentar el porcentaje a un 20%. Y aún más, si consideramos el conjunto de artículos, que no siendo entradas de ciencia y tecnología propiamente, incluyen en algunas de sus secciones contenidos científico-tecnológicos.

En cualquier caso, nuestro estudio evidencia el enorme volumen de información científica que Wikipedia pone al alcance de cualquier persona que disponga de conexión a la red. Un 10% de los artículos de la versión española, significa más de 100.000 artículos y, según nuestra estimación, un volumen cercano al millón de páginas.

Una tercera conclusión importante hace referencia a una de las cuestiones que más discusión social y mediática ha generado alrededor de la Wikipedia: la calidad y fiabilidad de su contenido. En nuestro estudio hemos utilizado algunas vías indirectas para evaluar dicha calidad – teniendo en cuenta la imposibilidad de evaluar directamente el contenido de una cantidad tan descomunal de artículos.

El estudio de las referencias que incluyen los artículos muestra que los editores de Wikipedia utilizan mayormente fuentes altamente fiables para escribirlos. En particular se apoyan a menudo en literatura científica primaria – y no en fuentes secundarias o divulgativas. La mayoría de las revistas científicas más importantes aparecen citadas o referenciadas en los artículos de Wikipedia y, lo que es aún más sorprendente, lo hacen siguiendo las frecuencias y distribución propias del mundo de la publicación científica, en la ciencia profesional. Los artículos, además, cuentan en general con numerosas referencias y, por lo tanto, permiten al lector profundizar en los contenidos si lo desea.

Además, la distribución temática de los artículos recoge un porcentaje mayoritario de los que pueden considerarse conceptos clave en las disciplinas que hemos analizado – a partir de las listas de conceptos o glosarios, de manuales universitarios estándar. Aunque nuestro estudio, y los propios editores, sugieren que aún queda camino por recorrer, en el sentido de mejorar la completar la cobertura temática en muchas disciplinas.

Curiosamente, la distribución temática se corresponde fuertemente con la distribución de la producción científica global en las distintas áreas de conocimiento. No sólo eso, sino que las relaciones entre estas áreas – analizadas a partir de los vínculos internos entre los artículos – muestran una gran semejanza con las relaciones entre disciplinas científicas que se dibujan a partir de los vínculos entre las publicaciones científicas. Algo que se hace evidente al comparar los mapas cognitivos de la ciencia con el mapa cognitivo que nosotros mismos hemos generado sobre Wikipedia.

Por todo ello, podemos concluir que la imagen de la ciencia que Wikipedia ofrece es, en resumen, una imagen muy ajustada de la ciencia contemporánea. El hecho de que Wikipedia esté editada – escrita – principalmente por personas que no son científicos profesionales, al margen de que pueda generar errores puntuales (y, en cualquier caso, no más que en las enciclopedias convencionales), no produce una imagen sesgada o deformada de la ciencia, ni una presencia de temas o perspectivas pseudocientíficas. Algo que también queda refrendado por nuestro análisis de las páginas de discusión.

Los editores, de hecho, muestran un perfil educativo considerablemente superior a la media de población general y en su mayoría tienen estudios universitarios. A pesar de no ser científicos profesionales, son capaces de manejar información y fuentes bibliográficas de primer nivel y de obtener de ellas los datos relevantes.

A pesar de la buena salud de la ciencia en la Wikipedia española – y teniendo en cuenta, como ya hemos dicho, la posibilidad y necesidad obvias de ampliar y mejorar el contenido existente – los científicos y científicas españolas están poco presentes en ella. Dado que muy pocos de ellos editan, las referencias a artículos científicos con autoría española no son muy frecuentes y, de hecho, las instituciones científicas españolas (universidades y centros de investigación, principalmente) aparecen poco referenciadas si lo comparamos con países cercanos del ámbito europeo. De hecho, sólo la mitad de los centros de investigación más importantes cuentan con una entrada propia en Wikipedia.

Por último, nuestro estudio pone de manifiesto el enorme potencial de Wikipedia en dos aspectos clave de la ciencia contemporánea. Por un lado, en la comunicación pública de la ciencia. Wikipedia es ya la principal fuente de información científica para la población y todo indica que lo seguirá siendo durante mucho tiempo, con todo lo que ello implica. Pero, además, nuestro estudio muestra la enorme capacidad de enrolar a los ciudadanos de forma activa en el mismo proceso de comunicación pública de la ciencia, y no sólo como meros receptores pasivos. En este sentido Wikipedia puede entenderse, también, como una plataforma sui generis de *ciencia ciudadana*.

En segundo lugar, Wikipedia se muestra como una herramienta muy útil para la propia enseñanza de la ciencia, en todos los niveles educativos, incluyendo el universitario. Los artículos de mayor calidad constituyen, a menudo, una excelente puerta de entrada o texto introductorio de gran utilidad para cualquier estudiante. Y, en el caso de los universitarios, les ofrece la posibilidad de explorar la literatura científica primaria más relevante en cada caso. Nuestros propios estudios anteriores, de hecho, muestran como la mayoría de profesores universitarios consultan con frecuencia Wikipedia – no tanto para temas de su propia especialidad sino para buscar información sobre ámbitos relacionados o más alejados de su experticia.

Este hecho, se complementa con la estrategia, que cada vez se utilizan más profesores universitarios en todo el mundo – aunque siguen siendo una minoría-, de diseñar actividades docentes en que los estudiantes, individualmente o por grupos, creen o mejoren artículos relacionado con la temática del curso. Se trata, en todo caso, de una de las posibles formas de implicar a la comunidad científica en la que sin duda es una de sus vías de comunicación más importantes con el resto de la sociedad.

8. Referencias

- Aibar, E. (2015). Wikipedia, Science, and Academia. In: Robert West, Leila Zia and Jure Leskovec (eds.). *Proceedings of the Workshop Wikipedia, a Social Media: Research Challenges and Opportunities*. ICWSM-15. Technical Report WS-15-19. Palo Alto (CA): AAAI Press; 2-5. Disponible en: <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10646>
- Aibar, E., Lladós, J., Minguiñón, J., Meseguer, A., & Lerga, M. (2015). Wikipedia at University: what Faculty Think and Do about it. *The Electronic Library* 33(4), pp. 668-683. Disponible en: <http://hdl.handle.net/10609/39442>
- Aigrain, P. (2012). *Sharing: Culture and Economy in the Internet Age*. Amsterdam: Amsterdam University Press. <http://www.sharing-thebook.com/content/download>
- Anderson, C. (2006). *The Long Tail: Why the Future of Business Is Selling Less of More*. New York: Hyperion.
- Benkler, Y. (2006). *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. New Haven, CT: Yale University Press.
- Brossard, D. & Scheufele, D.A. (2013). Science, new media, and the public. *Science*, 339(6115), pp. 40–41.
- Brown, A. R. (2011). Wikipedia As a Data Source for Political Scientists: Accuracy and Completeness of Coverage. *Political Science & Politics*, 44, pp. 339–343.
- Collins, H.M. y Pinch, T.J. 1993. *The Golem. What everyone should know about science*. Cambridge: Cambridge University Press.
- FECYT. (2012). *Encuesta de percepción social de la ciencia*. Disponible en: <http://www.fecyt.es/fecyt/docs/tmp/363174605.pdf>
- Ford, H., Sen, S., Musicant, D. & Miller, N. (2013). Getting to the Source: Where does Wikipedia Get Its Information From? *WikiSym '13*, August 5-7, Hong Kong, China.
- Giles, J. (2005). Internet Encyclopaedias Go Head To Head. *Nature*, 438, pp. 900–901.
- Groves, T., Figuerola, C.G., Alonso, J.L., & Quintanilla, M.A. (2014). Wikipedia as a tool for introducing social concerns into science education. En: Constantinou, C.P., Papadouris, N., and Hadjigeorgiou, A. (eds.). *Science Education Research For Evidence-based Teaching and Coherence in Learning (Proceedings of the ESERA 2013 Conference)*.
- Halfaker, A., and Taraborelli, D. (2015). Scholarly article citations in Wikipedia. *Figshare*. Disponible en: <http://dx.doi.org/10.6084/m9.figshare.1299540>
- Jemielniak, D. (2014). *Common Knowledge? An Ethnography of Wikipedia*. Stanford, CA: Stanford University Press.
- Luyt, B. & Tan, D. (2010). Improving Wikipedia's Credibility: References and Citations in a Sample of History Articles. *Journal of The American Society for Information Science and Technology*, 61(4):715–722.

Nielsen, F. (2007). Scientific Citations in Wikipedia. *First Monday*, 12 (8).

Mesgari, M., Okoli, C., Mehdi, M., Nielsen, F. AArup & Lanamäki, A. (2014). "The sum of all human knowledge": A systematic review of scholarly research on the content of Wikipedia. *Journal of the American Society for Information Science and Technology*. En: <http://spectrum.library.concordia.ca/978618/>

Moody, G. (2015). WikiGate' Raises Questions About Wikipedia's Commitment to Open Access. Jimmy Wales Asked to Reconsider Arrangement with Paywalled Science Publisher Elsevier. Disponible en: <http://arstechnica.com/science/2015/09/wikigate-raises-questions-about-wikipedias-commitment-to-open-access/>

National Science Board. (2012). *Science and Engineering Indicators 2012*. Washington, DC: National Science Foundation.

O'Sullivan, D. (2012). What Is An Encyclopedia? A Brief Historical Overview From Pliny To Wikipedia. In: Lovink, G. and Tkacz, N. (eds.). *Critical point of view: a Wikipedia Reader* (pp. 34-49). Amsterdam: Institute of Network Cultures.

Salah, A.A., Gao, C., Suchecki, K. & Scharnhorst, A. (2011). Generating Ambiguities: Mapping Category Names of Wikipedia to UDC Class Numbers. In: Lovink, G. and Tkacz, N. (eds.). *Critical point of view: a Wikipedia Reader* (pp. 34-49). Amsterdam: Institute of Network Cultures.

Teplitskiy, M., Grace, L. & Eamon, D. (2015). Amplifying the Impact of Open Access: Wikipedia and the Diffusion of Science. Forthcoming in: *Journal of the Association for Information Science and Technology*. <http://arxiv.org/abs/1506.07608v1>.

Voss, J. (2006). Collaborative thesaurus tagging the wikipedia way. Report, Wikimetrics. Disponible on-line: <http://arXiv.org/abs/cs/0604036>

Weber, S. (2004). *The success of open source*. Cambridge (MA): Harvard University Press.

Wedemeyer, B., Yakubova, N., Kallenbach, J., Ekdahl, A., Lesko, L., Reed, E., & Schwartz, K. (2008). Quality of the science articles on the English Wikipedia: Preliminary results. *Wikimania 2008*.