

El carácter enciclopédico de Wikipedia en español: aproximación a sus contenidos sobre Ciencias y Humanidades

J. Mario Fernández Montes¹, Miguel Ángel Iriarte Franco², José Emilio Mori Recio³ y Santiago Tejero Matía⁴

Recibido: 29 de noviembre de 2020 / Aceptado: 28 de marzo de 2021

Resumen. Partiendo de la dimensión enciclopédica de Wikipedia —marcada por la amplitud, variedad y ordenación sistemática de sus contenidos—, esta investigación busca identificar los campos de conocimiento presentes en ella y analizar cuantitativamente esa presencia. Con ese fin, selecciona una muestra de artículos —los destacados y buenos de Wikipedia en español— y, según agrupaciones de elaboración propia, los encuadra en los ámbitos STEM y HSSA, herederos de las clásicas Ciencias y Letras. Por medio de recursos de acceso a las bases de datos de Wikipedia y del tratamiento de dicha información, el trabajo estudia variables como el número de artículos, su extensión, la cantidad de visitas y el volumen de editores y ediciones. Las conclusiones ponen en relación la presencia de HSSA y STEM; la más destacada es el mayor volumen de artículos HSSA en cifras totales, que contrasta con el número más alto de visualizaciones medias de los STEM.

Palabras clave: Wikipedia; STEM; HSSA; Artículos Destacados; Artículos Buenos; Contenido Enciclopédico

[en] The Encyclopedic Character of Wikipedia in Spanish: An Approach to Its Contents on Sciences and Humanities

Abstract. Starting from the encyclopedic dimension of Wikipedia —marked by the breadth, variety and systematic arrangement of its contents—, this research seeks to identify the fields of knowledge present in it and analyze that presence quantitatively. To this end, it selects a sample of articles —the featured and good ones from Wikipedia in Spanish— and, according to groups of its own elaboration, frames them in the STEM and HSSA fields, inheritors of the classical Sciences and Arts. By means of access resources to Wikipedia databases and the subsequent treatment of said information, this project studies variables such as the number of articles, their length, the number of visits and the volume

-
- ¹ Universidad de Navarra (España)
E-mail: jmfernandez@unav.es
ORCID: <https://orcid.org/0000-0003-3352-0275>
 - ² Universidad de Navarra (España)
E-mail: mairiarte@unav.es
ORCID: <https://orcid.org/0000-0002-8692-7792>
 - ³ Wikipedia España
E-mail: joseemiliomori@wikimedia.es
ORCID: <https://orcid.org/0000-0003-1034-4598>
 - ⁴ Universidad de Navarra (España)
E-mail: stejero@unav.es
ORCID: <https://orcid.org/0000-0001-6177-6409>

of editors and editions. The conclusions show the relationship between HSSA and STEM; the most outstanding is the higher volume of HSSA articles in total figures, which contrasts with the higher number of average views of STEM.

Key words: Wikipedia; STEM; HSSA; Featured Articles; Good Articles; Encyclopedic Content

Sumario. 1. Introducción. 2. Objetivos. 3. Alcance y metodología. 4. Resultados y discusión. 5. Conclusiones. 6. Bibliografía.

Cómo citar. Fernández Montes, J. Mario; Iriarte Franco, Miguel Ángel; Mori Recio, José Emilio y Tejero Matía, Santiago (2021). El carácter enciclopédico de Wikipedia en español: aproximación a sus contenidos sobre Ciencias y Humanidades. *Área Abierta. Revista de comunicación audiovisual y publicitaria* 21 (2), 219-235, <https://dx.doi.org/10.5209/arab.72731>

1. Introducción

De los cinco pilares que, según sus responsables, sustentan Wikipedia⁵, suele referirse en primer lugar que se trata de “una enciclopedia que busca compendiar el conocimiento”⁶. En esto, sigue sin duda el ejemplo de la *Encyclopédie* dirigida por Denis Diderot —1713-1784— y Jean le Rond d’Alembert —1717-1783— entre 1751 y 1772, que contenía más de setenta y dos mil artículos elaborados por unos ciento cuarenta colaboradores, y cuya orientación universalista quedó reflejada en el significativo subtítulo de la obra: *Diccionario razonado de las ciencias, las artes y los oficios*.

En 2001, Wikipedia retomó aquella aspiración enciclopédica. Como expresó Jimmy Wales —uno de sus fundadores y actual presidente emérito de la Fundación Wikimedia— en una entrevista realizada por una revista española, “Wikipedia parte de una idea radical: dar acceso libre al conocimiento universal. Mi objetivo era —y es— ofrecer una enciclopedia gratuita y libre a todas las personas del planeta. Y en su idioma” (Uría, 2016). Actualmente, con más de cincuenta millones de artículos en trescientos trece idiomas, tres mil novecientos quince administradores y más de noventa y dos millones de usuarios, se ha convertido en un hito de gran relevancia en la historia del conocimiento.

Resulta indudable el interés de Wikipedia como objeto de estudio pues, como consecuencia de los indicadores antes citados, su página web es la quinta más visitada del mundo según datos de octubre de 2019, con una media de cinco mil cuatrocientos millones de visitas mensuales (Armstrong, 2019). Además, desde su nacimiento se han elaborado un buen número de artículos, libros y tesis doctorales que abordan diversos aspectos de su naturaleza y su evolución⁷.

⁵ Wikipedia. (2020). “Wikipedia: Los cinco pilares”. Recuperado de <https://w.wiki/z7b> (Fecha de acceso: 23/11/2020). En dicha web, se señalan estos rasgos fundamentales: 1) Wikipedia es una enciclopedia; 2) busca el “punto de vista neutral”; 3) es de contenido libre; 4) sigue unas normas de etiqueta; 5) no tiene normas firmes más allá de los cinco principios generales enunciados aquí.

⁶ Wikipedia. (2020). “Wikipedia: Lo que Wikipedia no es”. Recuperado de <https://w.wiki/zj7> (Fecha de acceso: 23/11/2020).

⁷ Wikipedia. (2020). “Anexo: Bibliografía sobre Wikipedia”. Recuperado de <https://w.wiki/zkZ> (Fecha de acceso: 21/11/2020).

En esa línea, la presente investigación analiza el carácter multidisciplinar de Wikipedia empleando para ello la división clásica de los saberes humanos, que fueron englobados desde el siglo XIX en las denominadas Ciencias y Letras y que en la actualidad han dado en llamarse STEM (Castro et al., 2020 y Hallinen, 2015) —Science, Technology, Engineering, Mathematics— y HSSA (Comisión Europea, 2020) —Humanities, Social Sciences, Arts—. Atrás quedaron otros intentos de organización del conocimiento como los estudios establecidos en el Imperio carolingio con los célebres Trivium —gramática, retórica y dialéctica— y Quadrivium —aritmética, astronomía, geometría y música— (Conrad, 2014) y, más adelante, los planes de trabajo de las universidades medievales y renacentistas.

Mientras las Humanidades atraviesan hoy momentos de incertidumbre y redefinición dentro de los sistemas educativos y en el ámbito laboral (Calderón, 2019), las disciplinas STEM en esos mismos sectores disfrutaban de una posición preponderante⁸. Sin embargo, voces provenientes del ámbito académico abogan por la integración de estos dos grandes campos. Un ejemplo es el profesor de Ciencias de la Computación de la Universidad de Kansas Lior Shamir: en un artículo reciente (Shamir, 2020), afirma que “integrar las humanidades con la ciencia y la ingeniería puede reanimar las ‘artes liberales’ y asegurar su existencia como disciplinas vitales para el bienestar de nuestra sociedad y nuestra cultura, ahora y en el futuro”.

En este contexto, comprobar si se da un correlato entre esa situación y la existencia en Wikipedia de artículos de un tipo u otro, su volumen y la actividad generada en torno a ellos, pareció una tarea pertinente, que llevaba, al menos, a las siguientes preguntas: ¿qué campos del saber están más presentes en Wikipedia? Y, al mismo tiempo, ¿cómo medir y objetivar esa presencia?

Conviene subrayar que, entre los posibles modos de afrontar esas cuestiones, el planteamiento de este trabajo es cuantitativo: trata de mostrar y analizar datos sobre aspectos numéricos concretos —que se especifican al explicar el propósito de la investigación—, mostrar relaciones entre esas informaciones y avanzar, en la medida de lo posible, hacia conclusiones que permitan señalar áreas de interés por parte de la comunidad de Wikipedia.

2. Objetivos

2.1. Objetivo principal

El objetivo de este artículo es el estudio cuantitativo de los contenidos de Wikipedia en español en torno a los dos grandes ámbitos de conocimiento en que se encuadran hoy nuestros saberes: STEM y HSSA.

Para ello, tiene en cuenta diversas variables que permiten hacerse cargo de la relevancia de un tema o un grupo de temas: el número total de artículos sobre él, la

⁸ Como señala Delgado (2019), “STEM es un área que continúa creciendo ya que los egresados de estos campos tienen alta demanda en el mercado laboral. El campo ha registrado un crecimiento del 17 % según el Departamento de Comercio en los Estados Unidos. Actualmente el 20 % de los trabajos están basados en las áreas STEM. Tan solo en Estados Unidos existen 26 millones de empleos en esa área y se estima que seguirán creciendo en los siguientes años. Además, los salarios de los egresados de carreras STEM generalmente son más altos que la media de todos los empleos”.

extensión de esos artículos, el número de visitas recibidas, la cantidad de editores y el número de ediciones realizadas.

Acotando así la finalidad del trabajo, quedan fuera de su ámbito dos cuestiones relacionadas con ella pero que requerirían análisis específicos: la calidad de los textos de Wikipedia y el perfil de sus usuarios. El primero de estos aspectos ha sido abordado, entre otros autores, por Raman et al. (2020), Shen, Salehi y Baldwin (2020), Pochampally y Karlapalem (2017) y Dang e Ignat (2016). Acerca del perfil de los usuarios, la propia Wikipedia explica sus distintos tipos⁹.

2.2. Otros objetivos

A partir del fin principal del presente estudio, y con la intención de concretar algunas líneas de análisis, los dos objetivos adicionales que complementan al central son los siguientes: a) conocer la presencia de los saberes STEM y HSSA en los artículos destacados y buenos de Wikipedia en español; y b) analizar las características de estos artículos: número, visualizaciones, tamaño, ediciones y editores.

3. Alcance y metodología

El punto de partida de este trabajo fue la estadística obtenida en Google Trends sobre búsquedas de ambos términos en los últimos diez años, que parecía una muestra razonablemente amplia. Allí, como refleja el Gráfico 1, se comprobó que las materias STEM son buscadas con más frecuencia. Aunque este asunto requeriría un análisis más detenido, ese dato confirmaba la intuición inicial de que las ciencias experimentales constituyen un ámbito de mayor envergadura tanto en el campo académico como entre los usuarios de la red en general.

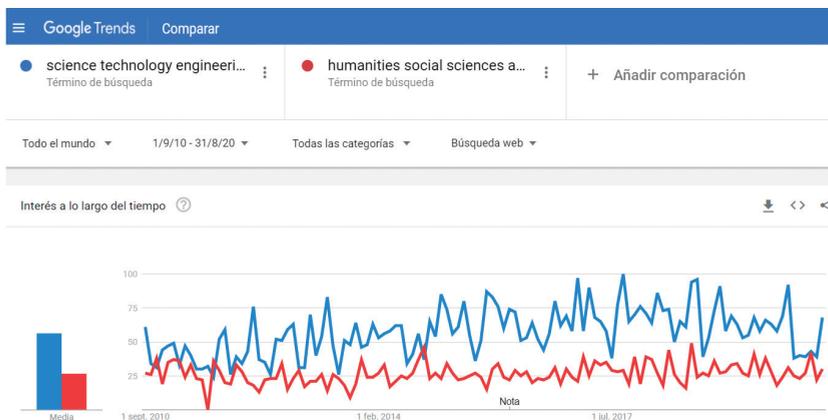


Gráfico 1. Búsqueda de términos “science technology engineering mathematics” y “humanities social sciences arts”. [Fuente: Google Trends]

⁹ Wikipedia. (2021): “Tipos de usuarios”. Recuperado de <https://w.wiki/zka> (Fecha de acceso: 12/01/2021).

Este indicador constituyó el arranque para tratar de comprobar hasta qué punto el interés mayoritario por las Ciencias en el motor de búsqueda más empleado del mundo tenía su reflejo o no en la principal enciclopedia *online*.

3.1. Alcance

Descartada por inabarcable una investigación de contenidos en varios idiomas, se ha optado por centrarse en el español por su proximidad y relevancia, pues el castellano es la novena lengua en Wikipedia según datos elaborados por la propia enciclopedia a partir del número de páginas de contenido. Y, dando un paso más, de todo el contenido de Wikipedia en español —1.641.778 artículos en noviembre de 2020— se ha seleccionado un total de tres mil novecientos veintisiete, de acuerdo con los siguientes tres criterios: 1) considerar solo artículos de Wikipedia en español, desde su inicio el veinte de mayo de 2001; 2) tomar los pertenecientes a los grupos Artículos Destacados y Artículos Buenos; y 3) utilizar las agrupaciones de contenido señaladas en los artículos destacados y buenos para la asignación a uno de los dos tipos de contenidos: HSSA y STEM.

Un artículo destacado es aquel que la comunidad de Wikipedia en español ha denominado así porque cumple una serie de requisitos: en concreto, tiene una excelente calidad al utilizar fuentes fiables, es verificable a través de fuentes externas merecedoras de confianza y se ofrece desde un punto de vista neutral. Además, cumple con las normas lingüísticas y ortográficas del castellano y del *Manual de estilo de Wikipedia*, y es completo, extenso y profundo. Para formar parte de los artículos destacados, cualquier usuario de Wikipedia puede añadirlo como candidato a destacado en la página de candidatos. Allí se someterá a votación revisando su estilo, integridad, precisión y neutralidad¹⁰.

Por su parte, un artículo bueno es aquel que, pese a cumplir los estándares de calidad de los artículos destacados, no pertenece a ese grupo por su brevedad. Junto a los requisitos ya mencionados, es estable, no enlaza a páginas de desambiguación y contiene imágenes con las licencias oportunas. Dichos artículos pueden no ser tan minuciosos y detallados como los destacados, pero no deben omitir ningún aspecto importante del tema. El proceso de designación para formar parte de los artículos buenos es más sencillo que el empleado con los destacados: cualquier usuario puede nominar un artículo si advierte que cumple con los criterios exigidos. En ese caso, tras la evaluación de un crítico imparcial y su aprobación —si corresponde—, se añadirá a la lista. Con todo, los artículos pueden ser retirados si se los promueve a destacados o si son removidos por no cumplir los criterios. Cualquier nominación, ya sea para aprobar un artículo o para eliminarlo de la lista, debe ir acompañada por argumentos en su discusión sobre cómo mejorarlos para alcanzar los criterios.

Como se ve, se ha optado por analizar una muestra de artículos de calidad reconocida por los wikipedistas: esta selección ha parecido la mejor manera de acotar el trabajo y de que pudiera llevar a conclusiones más representativas.

¹⁰ Wikipedia. (2020). “Wikipedia: Artículos destacados”. Recuperado de <https://w.wiki/zkp> (Fecha de acceso: 12/01/2021).

Por otro lado, se ha empleado el sistema de categorización utilizado por Wikipedia, ya que, como afirman Tramullas et al. (2017), sigue un esquema de clasificación universal desarrollado de forma colaborativa, semejante a un tesaurus (Archambault, Van De Belt, Grajales III et al., 2013).

3.2. Metodología

La investigación ha requerido una reflexión sobre los instrumentos más adecuados para el logro de sus objetivos y abundante tiempo para su consecución. Finalmente, tras valorar otras opciones, la obtención de la información y su análisis se ha realizado mediante los siguientes procedimientos y recursos:

1. Navegador de internet. Se ha empleado Google Chrome por ser el más utilizado entre los navegadores de internet (63,4 % de cuota de mercado según Statcounter GlobalStats, 2021).
2. Indización de artículos. Para los grupos de artículos destacados y buenos, se han usado los índices que los listan en ambas páginas, exportando el contenido a ficheros HTML y realizando posteriormente un tratamiento mediante fórmulas Excel.
3. Acceso a bases de datos. El acceso a las réplicas de las bases de datos de Wikipedia se ha llevado a cabo mediante las funcionalidades del software MediaWiki *Quarry beta*, utilizando comandos SQL y exportando los resultados a formato HTML. Las tareas anteriores se han complementado con la extracción de diversa información utilizando las órdenes *ad* y *ab* del bot Asimov desarrollado por el usuario “-jem-” (2020).
4. Descarga de estadísticas. Primero se ha realizado la obtención de los datos mediante las herramientas de Pageviews Analysis en Toolforge para la recogida de diversos valores de los artículos considerados: la herramienta solo proporciona información de diez artículos en cada consulta, por lo que desde la relación de páginas se han tenido que generar las URL correspondientes a todas ellas mediante fórmulas. A modo de ejemplo, esta sería la URL correspondiente a diez páginas de artículos destacados:
[https://pageviews.toolforge.org/?project=es.wikipedia.org&platform=all-access&agent=user&redirects=0&start=2015-07&end=2020-10&pages=Universidad_Nacional_Mayor_de_San_Marcos|Historia_de_Barcelona|México|Chile|Luz_en_la_pintura|Historia_de_la_ópera|Historia_del_arte|Racionalismo_\(arquitectura\)|Historia_del_desnudo_artístico|Club_Atlético_Rosario_Central](https://pageviews.toolforge.org/?project=es.wikipedia.org&platform=all-access&agent=user&redirects=0&start=2015-07&end=2020-10&pages=Universidad_Nacional_Mayor_de_San_Marcos|Historia_de_Barcelona|México|Chile|Luz_en_la_pintura|Historia_de_la_ópera|Historia_del_arte|Racionalismo_(arquitectura)|Historia_del_desnudo_artístico|Club_Atlético_Rosario_Central)
5. Herramientas de *web scrapping*. Extracción de la información de Toolforge de modo estructurado, utilizando una herramienta de *web scrapping* —en este caso, la extensión de Chrome Table Capture (GeorgeMike, 2020), por ser Chrome el navegador que dispone de más extensiones de uso libre—. Posteriormente los datos se han integrado en una tabla Excel.
6. *Clustering* de información no estructurada. Para el tratamiento de información no estructurada procedente de algunos ficheros en formato HTML, en primer lugar, se han utilizado técnicas de *clustering* para obtener variables binarias y, de esta forma, proceder posteriormente a su estudio con Excel mediante tablas dinámicas y fórmulas.

7. Normalización de valores. Con el fin de realizar algunas de las comparativas de los datos considerados, se ha procedido al cálculo de variables estandarizadas mediante una normalización de los valores en una escala de cero a uno.
8. Radar de contenidos. Para la representación gráfica de las variables normalizadas, se han utilizado representaciones gráficas en cuatro cuadrantes del tipo Gartner y Covey, con dos ejes X e Y con rango cero a uno, en los que se sitúan las variables en burbujas de tamaño proporcional a sus valores.

4. Resultados y discusión

4.1. Resultados

La exposición de los resultados obtenidos se va a realizar desde dos escenarios para procurar aportar una visión general y otra ponderada sobre la información obtenida: 1) aportación general de los valores absolutos de cada una de las variables por separado; y 2) normalización de los valores para explicar su aportación en función de las tres agrupaciones generales que se definen.

4.1.1. Aportación general de los valores absolutos de cada una de las variables por separado

En conexión con uno de los objetivos del estudio (2.2.b), se han considerado las siguientes variables: número de artículos, número de visualizaciones, número de ediciones, número de editores, tamaño, grupo al que pertenece —destacados o buenos— y tipo —HSSA o STEM—.

Tabla 1. Distribución de contenidos destacados y buenos. [Fuente: elaboración propia].

Destacados HSSA	Destacados STEM	Buenos HSSA	Buenos STEM
Arte y arquitectura	Alimentación	Arte	Ciencias naturales
Ciencias sociales	Astronomía y astrofísica	Ciencias sociales	Ingeniería
Cine, fotografía, televisión y entretenimiento	Biología y medicina	Deportes	Matemáticas y lógica
Cultura y sociedad	Física, química y matemáticas	Humanidades	
Deportes	Medio ambiente	Vida social	
Derecho	Tecnología e informática		
Filosofía			
Fuerzas armadas y guerra			
Geografía y lugares			
Historia			
Lenguaje			
Literatura y teatro			
Música			
Países			
Política y gobierno			
Prehistoria			
Religión y mitología			

La asignación de cada artículo a su correspondiente tipo HSSA o STEM se ha realizado distribuyéndolos en uno de los dos tipos según muestra la Tabla 1. En total se han asignado treinta y un subgrupos, veintitrés de artículos destacados —diecisiete HSSA, seis STEM— y ocho de buenos —cinco HSSA y tres STEM—.

En la Tabla 2 se presentan los resultados numéricos obtenidos y se explican a continuación para cada variable.

Tabla 2. Distribución de los valores de número de artículos, visualizaciones, ediciones y editores por grupo y tipo. [Fuente: elaboración propia].

	Grupo	HSSA	STEM	TOTAL	Ratio HSSA/STEM
NÚMERO DE ARTÍCULOS	Destacados	942	197	1.139	4,7
	Buenos	2.292	496	2.788	
	Total	3.234	693	3.927	
VISUALIZACIONES	Destacados	1.036.149.866	258.889.603	1.295.039.469	3,7
	Buenos	942.578.370	273.363.888	1.215.942.258	
	Total	1.978.728.236	532.253.491	2.510.981.727	
	Media/artículo	611.852	768.043		
TAMAÑO	Destacados	91.614.725	16.160.483	107.775.208	5,7
	Buenos	118.840.197	20.846.698	139.686.895	
	Total	210.454.922	37.007.181	247.462.103	
	Media/artículo	65.076	53.401		
EDICIONES	Destacados	298.602	50.148	348.750	6,1
	Buenos	344.382	55.512	399.894	
	Total	642.984	105.660	748.644	
	Media/artículo	199	152		
EDITORES	Destacados	111.366	22.114	133.480	5,2
	Buenos	131.563	24.733	156.296	
	Total	242.929	46.847	289.776	
	Media/artículo	75,1	67,6		

4.1.1.1. Número de artículos

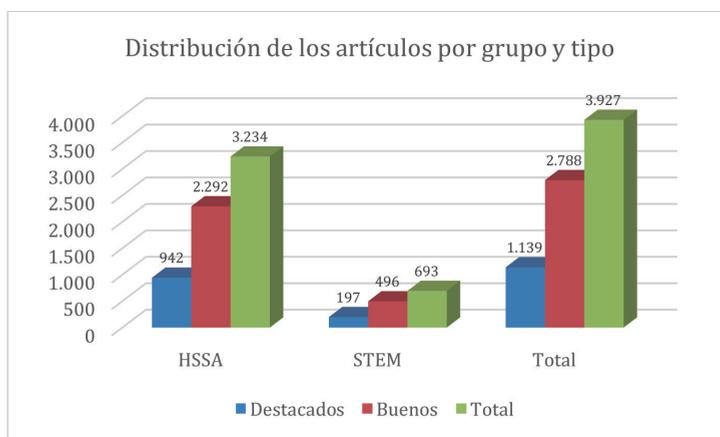


Gráfico 2. Distribución de los artículos por grupo y tipo. [Fuente: elaboración propia].

En la Tabla 2 se recoge la distribución del total de los tres mil novecientos diecisiete artículos seleccionados, en la que se puede apreciar que, en volumen absoluto de número de artículos, HSSA tiene 4,7 veces más que STEM.

4.1.1.2. Número de visualizaciones

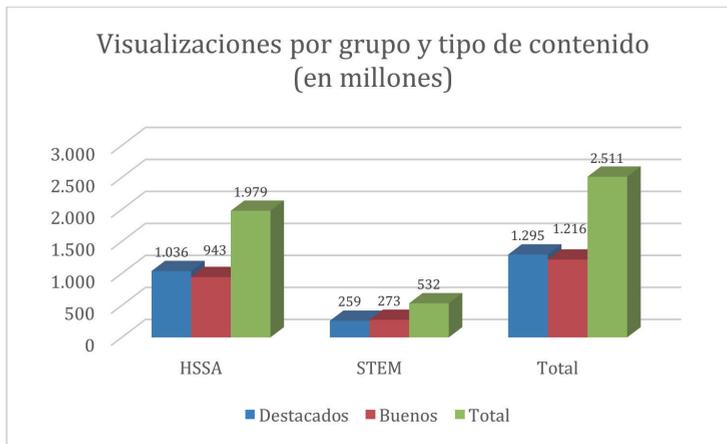


Gráfico 3. Total de visualizaciones por grupo y tipo de contenido. [Fuente: elaboración propia].

En la Tabla 2 se muestra el total de las 2.510.981.727 visualizaciones que han recibido los artículos desde julio de 2015 a octubre de 2020, ambos meses incluidos. Se ha seleccionado este periodo para todos los artículos, ya que con anterioridad a esta fecha los recursos de *Toolforge* no proporcionan datos. El campo en el que se localizaba esta información fue eliminado a partir de la versión MediaWiki 1.25, debido a que no solo contabilizaba usuarios que accedían a las páginas, sino que también incluía todos los *bots* que las visitaban (MediaWiki, 2020).

Como puede apreciarse de nuevo, el número de visualizaciones de HSSA supera al de STEM —es 3,7 veces mayor—.

4.1.1.3. Tamaño

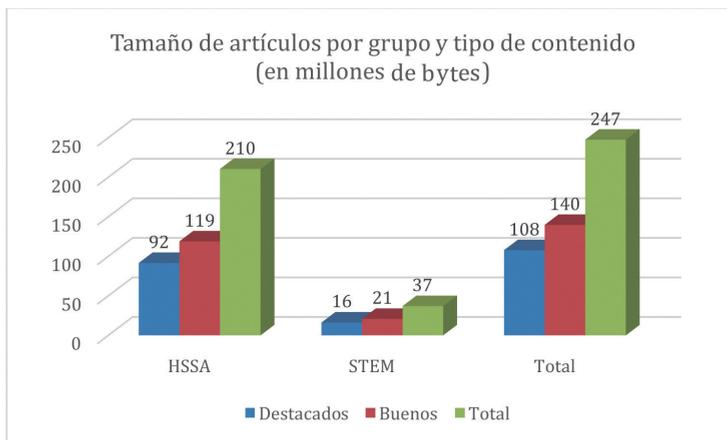


Gráfico 4. Tamaño (en millones de *bytes*) de artículos por grupo y tipo de contenido. [Fuente: elaboración propia].

En la Tabla 2 se muestra el tamaño —en *bytes*— de todos los artículos, con un total de 247.462.103 bytes. En este caso, la ratio de tamaños totales entre HSSA y STEM resulta ser de 5,7.

4.1.1.4. Número de ediciones

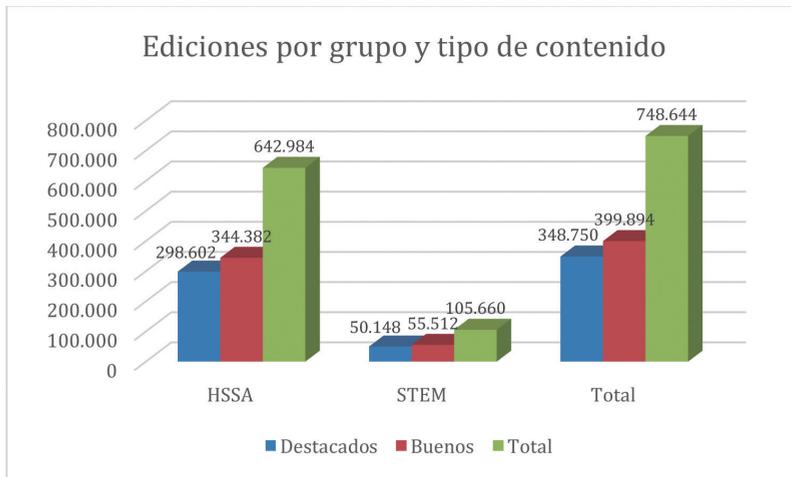


Gráfico 5. Número de ediciones por grupo y tipo de contenido.
[Fuente: elaboración propia].

La Tabla 2 presenta el número de ediciones de todos los artículos, con un total de 748.644. Nuevamente, las ediciones de HSSA son superiores y suponen 6,1 veces las de STEM.

4.1.1.5. Número de editores

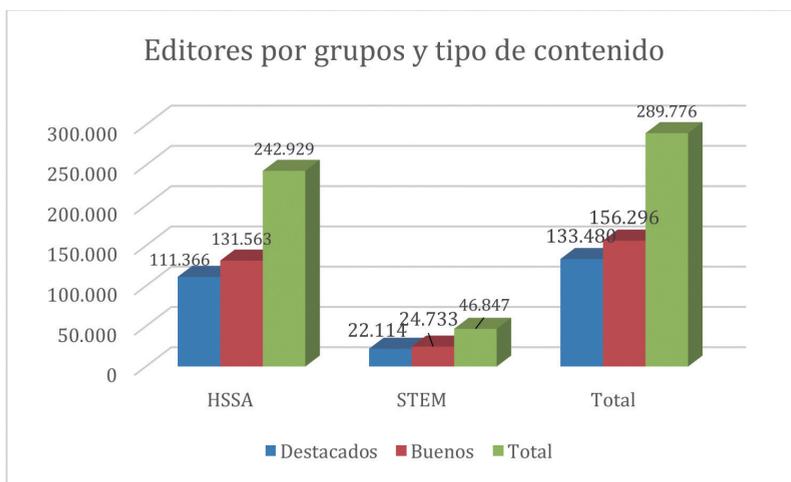


Gráfico 6: Número de editores por grupos y tipo de contenido.
[Fuente: elaboración propia].

Sobre el número de editores, la información que aporta Toolforge es la cifra total de editores que han editado un artículo: alguno de ellos ha podido editar también otro artículo, en el que sería contabilizado también como editor. Hay que hacer esta observación ya que el número total de editores distintos que han editado estos artículos es menor que la suma del total de editores que se utiliza en los datos.

De esta forma, el número total de editores que han editado estos artículos es de 289.776, como muestra la Tabla 2. De igual modo, la ratio entre editores de HSSA y de STEM es de 5,2.

4.1.2. Normalización de los valores para explicar su aportación en función de las tres agrupaciones generales que se definen

Como se ha visto en los datos del apartado previo, la gran diferencia entre el número de artículos de los grupos destacados y buenos —como se refería anteriormente, los buenos son 4,6 veces más que los destacados— plantea la necesidad de que se normalicen los valores de las variables consideradas en un índice que elimine la distorsión introducida por la divergencia de los valores absolutos.

Para ello, se han normalizado los valores en una escala de cero a uno para cada una de las tres agrupaciones que se consideran (Tabla 3):

Tabla 3. Agrupaciones generales de artículos.
[Fuente: elaboración propia].

Agrupaciones generales artículos	
Agrupación 1: según categoría	
G1-D	Destacados
G1-B	Buenos
Agrupación 2: según tipo de contenido	
G2-H	HSSA
G2-S	STEM
Agrupación 3: según categoría y tipo de contenido	
G3-DH	Destacados-HSSA
G3-DS	Destacados-STEM
G3-BH	Buenos-HSSA
G3-BS	Buenos-STEM

Se calcula este índice para las variables denominadas *tamaño*, *visualizaciones*, *ediciones* y *editores* y se ponen en relación entre sí en un solo gráfico de cuadrantes.

Para ofrecer una visión de conjunto de algunas de las variables consideradas, se ha elaborado una matriz de cuatro cuadrantes con tres variables: el tamaño de la burbuja, que expresa el número de artículos; y los ejes X e Y, que recogen los valores de las distintas variables normalizados entre cero y uno. Esta matriz, denominada *Radar de Contenidos*, pretende ofrecer una radiografía de la relación numérica entre las variables.

A modo de ejemplo, a continuación, se añaden las claves que ayudan a interpretar la matriz adecuadamente para las variables *tamaño* y *visualizaciones*:

Cuadrante I —superior izquierda—: en él se concentran artículos que, teniendo tamaño pequeño, consiguen un alto número de visualizaciones.

Cuadrante II —superior derecha—: en él se encuentran los artículos con mayor tamaño y más visualizaciones.

Cuadrante III: —inferior izquierda—: en él aparecen los artículos con menor tamaño y menos visualizaciones.

Cuadrante IV —inferior derecha—: reúne los artículos con mayor tamaño y menores visualizaciones.

Seguidamente se detallan los resultados para las tres agrupaciones de artículos.

4.1.2.1. Agrupación 1: según grupo de Destacados y Buenos

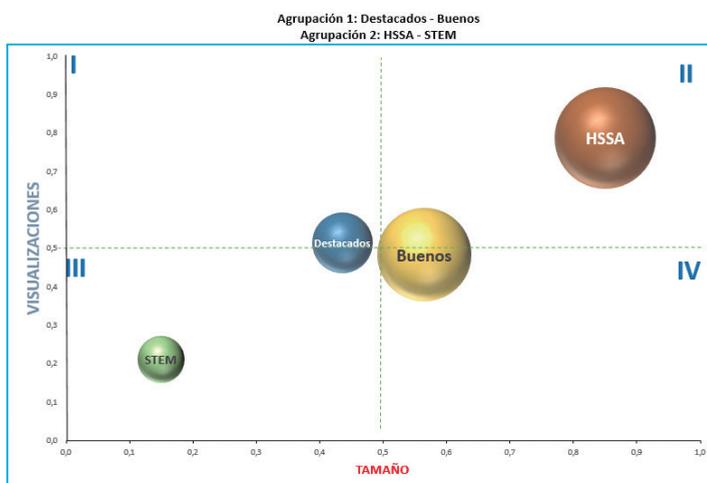


Gráfico 7. Radar de Contenidos para los índices por separado de Grupo y Tipo de contenido. Variables: tamaño y visualizaciones. [Fuente: elaboración propia].

En el Gráfico 7 se observa en general un posicionamiento ligeramente menor en el índice de Tamaño por grupo del grupo Destacados con un valor de 0,44 frente al valor de 0,56 de Buenos.

En Visualizaciones sigue dándose un equilibrio entre ambos grupos, con valores de 0,52 para *Destacados* y 0,48 para Buenos.

El tamaño de la burbuja, que en este caso marca el índice del número de artículos en cada grupo, es notablemente mayor en Buenos —0,71— que en Destacados —0,29—.

4.1.2.2. Agrupación 2: según contenido

En cuanto al tipo de contenido, en el Gráfico 7 se observa que el índice de tamaño del tipo HSSA es claramente superior al situarse en la parte alta del cuadrante II con

un valor de 0,85, frente al de STEM, que se localiza en la parte central del cuadrante III con un 0,15.

En el índice de visualizaciones se dan resultados similares en la ubicación por cuadrantes, con valores de 0,79 para HSSA y 0,21 para el tipo STEM.

Sobre el tamaño de la burbuja, que en este caso muestra el índice de número de artículos en cada tipo de contenido, sigue siendo muy superior en el tipo HSSA —0,82— frente al tipo STEM —0,18—.

4.1.2.3. Agrupación 3: según grupo y tipo de contenido

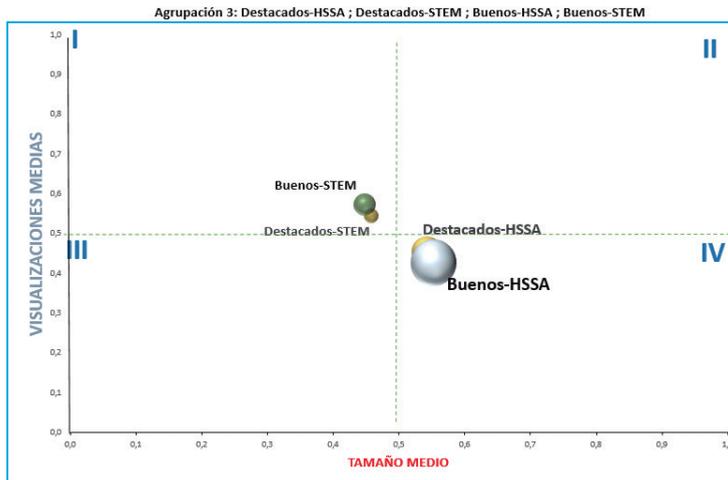


Gráfico 8. Radar de Contenidos para los índices conjuntos de agrupaciones y tipo de contenido. Variables: tamaño medio y visualizaciones medias. [Fuente: elaboración propia].

En el Gráfico 8 se han considerado los índices conjuntos de las cuatro agrupaciones que se derivan de la combinación de grupos y tipo de contenido, y en los ejes X e Y se representan los índices de tamaño medio y visualizaciones medias para dichas agrupaciones.

En cuanto al índice de tamaño medio, se observa un valor ligeramente superior en la agrupación Destacados-HSSA, con un índice de 0,54, frente a la de Destacados-STEM con un índice de 0,46. En cambio, en las otras dos agrupaciones se comprueba un mejor posicionamiento de la agrupación Buenos-HSSA, con un índice de 0,55 frente a la agrupación Buenos-STEM, con un 0,45.

Por lo que se refiere a los valores de visualizaciones medias, la agrupación Destacados-HSSA, con un índice de 0,46, está ligeramente por debajo de Destacados-STEM, que presenta un índice de 0,54. Para las otras agrupaciones, la de Buenos-HSSA, con un índice de 0,43, se posiciona por debajo de la de Buenos-STEM —0,57—.

El tamaño de la burbuja en este apartado muestra el índice del número de artículos medios en cada agrupación: Destacados-HSSA ocupa el segundo lugar con un valor de 0,24; Destacados-STEM es el más bajo, con un 0,05, mientras Buenos-HSSA es el mayor, con un valor de 0,58 y Buenos-STEM ocupa el tercer lugar, con un valor de 0,13.

Así se puede deducir que, teniendo un tamaño medio equivalente y aun habiendo más artículos Destacados-HSSA, sin embargo, los artículos Destacados-STEM son más visitados que el resto de agrupaciones.

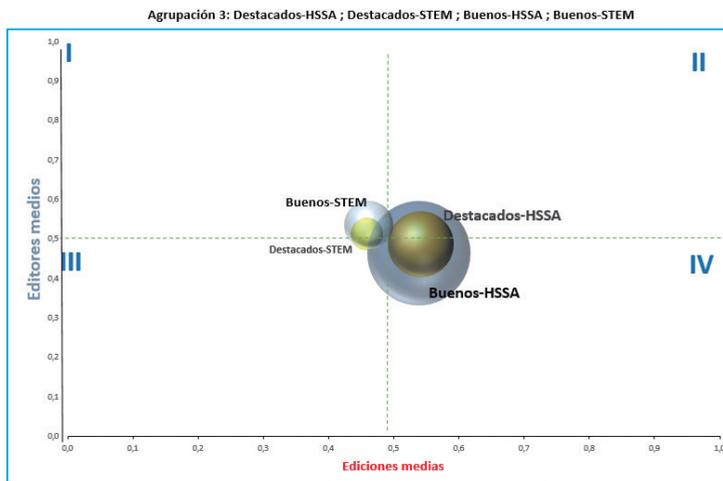


Gráfico 9: Radar de Contenidos para los índices conjuntos de agrupaciones y tipo de contenido. Variables: ediciones medias y editores medios. [Fuente: elaboración propia].

En el Gráfico 9 se han considerado los índices conjuntos de las cuatro agrupaciones que se derivan de la combinación de grupos y tipo de contenido, y en los ejes X e Y se representan los índices de ediciones medias y editores medios para dichas agrupaciones.

En cuanto al índice de ediciones medias, se observa un valor ligeramente superior en las agrupaciones Destacados-HSSA, con un índice de 0,54 frente al de Destacados-STEM (0,46). Pero en las otras dos agrupaciones se ve un mejor posicionamiento de la de Buenos-HSSA, con un índice de 0,54, frente a la de Buenos-STEM, con un índice de 0,46.

Al fijarse en el índice de editores medios, se observa un equilibrio de la agrupación Destacados-HSSA, con un índice de 0,49, y de Destacados-STEM —0,51—. Para las otras agrupaciones, la de Buenos-HSSA, con un índice de 0,46, se sitúa por debajo de la de Buenos-STEM —0,54—.

El tamaño de la burbuja en este caso es el ya referido para el Gráfico 8.

Esto puede indicar que los editores tienen comportamientos similares en las cuatro agrupaciones consideradas a la hora de escribir contenidos de artículos, ya que no hay ninguna que sea especialmente destacada, como se observa en el Gráfico 9, donde todos los valores se sitúan muy cercanos al punto de intersección central de los cuatro cuadrantes. Y lo mismo puede decirse de las ediciones.

4.2. Discusión

Tras mostrar las tablas y gráficas anteriores, y antes de pasar a las conclusiones del trabajo, parece oportuno mencionar algunas líneas de posible análisis de lo expuesto

hasta ahora. Con ello se busca aportar una cierta interpretación de los datos y, por otra parte, plantear algunas hipótesis de interés para el lector. Estudios como este no pretenden ni pueden ser excesivamente audaces ni realizar grandes declaraciones o descubrimientos sobre Wikipedia en su conjunto —algo que seguramente no llevaría a resultados prácticos—. Sin embargo, centrarse en estudiar con detalle y criterios adecuados determinadas cuestiones concretas, como los contenidos STEM y HSSA, puede contribuir a la reflexión y a desarrollar acciones igualmente concretas dentro de los proyectos de la comunidad de Wikipedia.

Según muestra Google Trends, en los últimos años los contenidos STEM son más buscados en Google que los HSSA. Sin embargo, según ha mostrado este trabajo, entre los artículos destacados y buenos en Wikipedia en español, en números absolutos, no se verifica esa tendencia. Este hecho podría llevar a varias preguntas sobre el perfil de los editores y los usuarios de Wikipedia. Así, podría pensarse, por ejemplo, que los textos científicos, por su mayor complejidad, sobre todo si requieren un respaldo experimental, tienden a aparecer en publicaciones de mayor especialización que Wikipedia. Eso explicaría que el número de editores y de visualizaciones de contenidos STEM en Wikipedia sea más bajo que en el caso de los HSSA.

Dando un paso más, cabría plantearse si Wikipedia —al menos en español— está contribuyendo a evitar una brecha entre el conocimiento de nivel científico profesional y otro con calidad contrastada, pero de alcance menos especializado. El análisis de los datos de este trabajo lleva a pensar que es así, en la medida que el predominio de los contenidos HSSA de Wikipedia amortigua la tendencia general a los STEM mostrada por Google, que se orientan a una mayor especialización.

Por otra parte, la naturaleza de los dos tipos de materias puede explicar en buena medida la cuestión de la extensión y las visualizaciones de los artículos, ya que en HSSA resulta fundamental el empleo de abundante información y datos que generalmente no se pueden memorizar, mientras que en STEM es más frecuente manejar leyes generales que luego se aplican en distintos campos; esas leyes, teoremas y análisis científicos requieren una demostración más o menos formal, exigen dedicar más tiempo y trabajo y se destinan con mayor frecuencia a las publicaciones científicas y especializadas.

En todo caso, nos movemos en el ámbito de ideas o conjeturas de trabajo que requerirían una mayor profundización, en primer lugar, con investigaciones que abarcaran todos los artículos de Wikipedia en español, y, en un momento posterior, en el ámbito angloparlante, cuyo peso en Wikipedia es claramente mayor que el de otras comunidades lingüísticas.

5. Conclusiones

En este apartado se exponen las ideas finales del trabajo, resultado de la elaboración de la información obtenida y de la reflexión sobre ella. Se comienza por dos más generales para luego señalar otras más específicas. Se concluye con la sugerencia de un posible campo que, tras la investigación realizada, parece merecedor de atención y de otros estudios futuros.

1. El presente análisis, especialmente con datos aportados en su primer epígrafe, ha confirmado el carácter enciclopédico que Wikipedia se atribuye. Esa aspi-

ración de sus fundadores se ve refrendada por la cantidad y la variedad de la información que ofrece a sus millones de usuarios en todo el mundo. Esto hace que pueda decirse que Wikipedia llega a su vigésimo aniversario con una situación asentada entre las fuentes de conocimiento más empleadas en nuestros días.

2. Una valoración de conjunto permite afirmar que se ha abordado satisfactoriamente el objetivo general de llevar a cabo un análisis cuantitativo de los contenidos STEM y HSSA en Wikipedia en español.
3. La agrupación de artículos en las tipologías STEM y HSSA ha resultado muy útil para estudiar la diversidad de contenido de los artículos de Wikipedia considerados. Además, la tarea de agrupación de artículos desarrollada por Wikipedia encaja con precisión suficiente con estas dos grandes áreas de conocimiento.
4. En cifras absolutas, tanto el número de artículos como el de visualizaciones y el tamaño —bytes— es mayor en los artículos del tipo HSSA que en los de STEM. Por contraste, en valores relativos los artículos de STEM son más visualizados que los de HSSA.
5. Los editores tienen un comportamiento similar a la hora de realizar las ediciones de los artículos de cualquiera de las agrupaciones, habiendo homogeneidad en el número medio de editores, incluso considerando la mayor presencia de artículos buenos frente a los destacados.
6. Para acercarse más a la realidad de Wikipedia en estos aspectos, sería conveniente analizar con las agrupaciones empleadas en este trabajo el ámbito angloparlante, por su mayor volumen de contenidos, o cualquier otro proyecto de la enciclopedia en otro idioma.

6. Bibliografía

- Archambault, P. M.; Van De Belt, T. H., Grajales III, F. J. et al., (2013). “Wikis and collaborative writing applications in health care: A scoping review”, *Journal of Medical Internet Research*, vol. XV, núm. 10, DOI: <https://doi.org/10.2196/jmir.2787>
- Armstrong, M. (2019). “The World’s Most Popular Websites”. Recuperado de <https://www.statista.com/chart/17613/most-popular-websites/> (Fecha de acceso: 21/11/2020).
- Calderón, R. (2019). “¡Prohibamos las humanidades!”. *Telos*, núm. 112, 10-15. <https://telos.fundaciontelefonica.com/telos-112-humanidades-en-un-mundo-stema-autor-invitado-reyes-caledon-prohibamos-las-humanidades/> (Fecha de acceso: 15/11/2020).
- Castro Inostroza, Á., Iturbe Sarunic, C., Jiménez Villarroel, R., Silva Hormázabal, M. (2020). “Stem or humanities education? “A reflection on the integral formation of the 21st century citizen”, *Utopía y Praxis Latinoamericana*, vol. XXV, núm. Extra 9, 197-208. DOI: <https://doi.org/10.5281/zenodo.4110904>
- Comisión Europea. (2020). “Horizon 2020: Social Sciences and Humanities”. Recuperado de <https://ec.europa.eu/programmes/horizon2020/en/area/social-sciences-and-humanities> (Fecha de acceso: 21/11/2020).
- Conrad, Leon. (2014). “Integration and the liberal arts: A historical overview”. *On the Horizon*, vol. 22, núm. 1, 46-56 <https://doi.org/10.1108/OTH-11-2013-0045>
- Dang, Quang-Vinh, Ignat, Claudia Lavinia. (2016). “Measuring quality of collaboratively edited documents: The case of Wikipedia”, en *IEEE 2nd International Conference on Col-*

- laboration and Internet Computing, IEEE CIC 20166 January 2017*, pp. 266-275, DOI: <https://doi.org/10.1109/CIC.2016.42>
- Delgado, Paulette. (2019). “Educación STEM: ¿qué es y cómo sacarle provecho? Recuperado de <https://observatorio.tec.mx/edu-news/educacion-stem-que-es-y-como-sacarle-provecho> (Fecha de acceso: 21/11/2020).
- Diderot, Denis. (1974). *La Enciclopedia*: (selección), Madrid: Guadarrama.
- Diderot, Denis, Le Rond d’Alembert, Jean. (1966). *Encyclopedie, ou dictionnaire raisonne des sciences, des arts et des metiers par une societ  de gens de lettres*, Stuttgart-Bad Cannstatt: Friedrich Frommann
- GeorgeMike.com. (2020). “Table Capture”. Recuperado de <https://chrome.google.com/webstore/detail/table-capture/iebjdmgckacobjpjhpcplhebcmeop> (Fecha de acceso: 22/11/2020).
- Hallinen, J. (2015). “STEM Education Curriculum”. Encyclop dia Britannica, Recuperado de <https://www.britannica.com/topic/STEM-education> (Fecha de acceso: 21/11/2020).
- Jem. (2020). “Asimov: Informaci n general”. Recuperado de <https://jembot.toolforge.org/asimovi.php> (Fecha de acceso: 23/11/2020).
- MediaWiki. (2020). “MediaWiki 1.25”. Recuperado de <https://w.wiki/zkw> (Fecha de acceso: 22/11/2020).
- MusikAnimal; Kaldari; Ruiz-Forns, Marcel (2020). “Pagesviews analysis”. *Wikimedia Toolforge*. Recuperado de <https://tools.wmflabs.org/admin/tool/pageviews> (Fecha de acceso: 22/11/2020).
- Pochampally, Y.; Karlapalem, K. (2017). “Notability determination for Wikipedia”, En *26th International World Wide Web Conference 2017*, WWW 2017 Companion 2019, pp. 1641-1646, DOI: <https://doi.org/10.1145/3041021.3053361>
- Raman, N.; Sauerberg, N.; Fisher, J.; Narayan, S. (2020). “Classifying Wikipedia Article Quality with Revision History Networks”. En *OpenSym 2020: Proceedings of the 16th International Symposium on Open Collaboration*. Nueva York: Association for Computing Machinery - TJEF: The John Ernest Foundation. <https://doi.org/10.1145/3412569.3412575>
- Shamir, L. (2020). “A Case Against the STEM Rush”. Recuperado de <https://www.insidehighered.com/views/2020/02/03/computer-scientist-urges-more-support-humanities-opinion> (Fecha de acceso, 14/01/2021).
- Shen, A.; Salehi, B.; Qi, J.; Baldwin, T. (2020). “A multimodal approach to assessing document quality”, *Journal of Artificial Intelligence Research*, vol. LXVIII, pp. 607-632. DOI: <https://doi.org/10.1126/science.aa0462>
- Statcounter GlobalStats. Browser Market Share Worldwide. Recuperado de <https://gs.statcounter.com/browser-market-share#monthly-202001-202012> (Fecha de acceso: 9/01/2021).
- Tramullas, J.; S nchez-Casab n, A.; Garrido-Picazo, P. (2017). “Utilizaci n de categor as de Wikipedia en procesos de organizaci n de informaci n. Hacia una revisi n cualitativa”. En el *III Congresso ISKO Espanha e Portugal, XIII Congresso ISKO Espanha Perspectivas de investiga o em representa o e organiza o do conhecimento: atualidade e tend ncias*, pp. 601-609. Recuperado de: http://eprints.rclis.org/39931/1/isko2017_602_610.pdf (Fecha de acceso: 20/11/2020).
- Ur a, I. (2016). “Jimmy Wales. Wikipedia, el conocimiento democr tico”. *Nuestro Tiempo*, vol. LXI, n m. 690, 6-17.

