Desambiguación en *Wikipedia*: exploración de los mecanismos de control de autoridades en la enciclopedia colaborativa

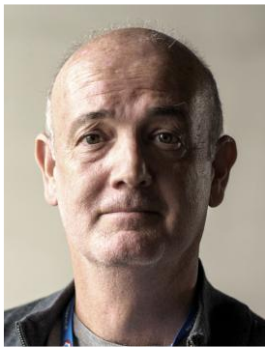
Disambiguation in *Wikipedia*: exploring the mechanisms of authority control in the collaborative encyclopedia

Florencia Claes; Tomás Saorín

Cómo citar este artículo:

Claes, Florencia; **Saorín, Tomás** (2025). "Desambiguación en *Wikipedia*: exploración de los mecanismos de control de autoridades en la enciclopedia colaborativa [Disambiguation in *Wikipedia*: exploring the mechanisms of authority control in the collaborative encyclopedia]". *Infonomy*, 3(3) e25017. *https://doi.org/10.3145/infonomy.25.017*

Florencia Claes
https://orcid.org/0000-0003-4291-0507
https://directorioexit.info/ficha3580
Universidad Rey Juan Calos
Facultad de Ciencias de la Comunicación
Camino del Molino, 5
28943 Madrid, España
florencia.claes@urjc.es



Tomás Saorín
https://orcid.org/0000-0001-9448-0866
https://directorioexit.info/ficha1039
Universidad de Murcia
Facultad de Comunicación y Documentación
Campus de Espinardo
30071 Murcia, España
tsp@um.es

CC BY

Resumen
Se estudia el sistema de desambiguación de *Wikipedia* desde la perspectiva de los instrumentos de control de autoridades y los vocabularios controlados. *Wikipedia* es un sistema de enciclopedias en diferentes idiomas, organizado en entradas conectadas y referencias cruzadas. Este trabajo explora varios aspectos de las páginas de desambiguación, delimitando su nivel de uso y ratios en las ediciones más extensas. También se exploran otras características relacionadas con el nivel de actividad de sus editores, con el total de artículos afectados por la desambiguación, su tipología, la equivalencia interlingüística y su modelización como datos estructurados. Se presentan, además, datos y métricas para el castellano y el catalán.

Palabras clave
Control de autoridades; Vocabularios controlados; *Wikipedia*; Organización del conocimiento; Desambiguación; Enciclopedia; Datos estructurados.

Abstract
This paper studies *Wikipedia*'s disambiguation system from the perspective of authority control tools and controlled vocabularies. *Wikipedia* is a system of encyclopedias in different languages, organized into connected entries and cross-references. This paper explores various aspects of disambiguation pages, delineating their level of use and ratios in the most extensive edits. It also explores other characteristics related to the level of activity of its editors, the total number of articles affected by disambiguation, its typology, interlingual equivalence, and its modeling as structured data. Data and metrics are also presented for Castilian and Catalan.

Keywords
Authority control; Controlled vocabularies; *Wikipedia*; Knowledge organization; Disambiguation; Encyclopedia; Structured data.

1. Introducción

En este trabajo presentamos una sección poco conocida de la enciclopedia libre colaborativa *Wikipedia*: sus páginas de desambiguación. El trabajo más visible de quienes editan voluntariamente es el de la elaboración y mejora de artículos enciclopédicos, escritos desde un punto de vista neutral y apoyados en fuentes verificables y de calidad. Además de ese trabajo en el contenido, existen otras numerosas tareas que podríamos denominar genéricamente de mantenimiento y organización del contenido, que son muy importantes para que el proyecto funcione correctamente. De entre esos otros "trabajos enciclopédicos" vamos a centrarnos en los que se conocen en *Wikipedia* como desambiguación.

¿Qué son las páginas de desambiguación? Dado que en la enciclopedia el punto de acceso principal al contenido son los títulos de los artículos, estás páginas recopilan títulos parecidos entre los que puede haber confusión, y enlazan a todos los diversos artículos añadiendo una concisa explicación del contexto y uso. En las siguientes figuras de ejemplo se ve claro el objeto sobre el que vamos a tratar:

Figura 1. Página de desambiguación en español para el término *Once*
https://es.wikipedia.org/wiki/Once_(desambiguaci%C3%B3n)

Figura 2. Página de desambiguación en español para el término *Taco*
https://es.wikipedia.org/wiki/Taco_(desambiguaci%C3%B3n)

En el contexto de las bibliotecas y archivos estas páginas se entenderían desde las áreas de control de puntos de acceso, el control de autoridades y el control de vocabulario relacionadas con la catalogación y la descripción (**Joudrey**; **Taylor**, 2018). Las tareas de diferenciar entre significados también forman parte de las técnicas lexicográficas de larga tradición para producir diccionarios y terminologías. En un sentido muy amplio podríamos hablar de instrumentos o sistemas para la organización del conocimiento (*Knowledge Organization Systems, KOS*), en los que siempre está presente la necesidad de diferenciar entre términos y conceptos similares. En los siguientes apartados realizaremos algunas consideraciones esquemáticas sobre aspectos de organización del conocimiento presentes en las enciclopedias, diccionarios, vocabularios controlados, bases de datos de autoridades y diccionarios.

2. ¿Es la enciclopedia un sistema de organización del conocimiento?

Una enciclopedia es un producto intermedio, tipificado por algunas escuelas de pensamiento como documento "terciario", que incorpora elementos de los documentos primarios (el contenido, el texto) y de los secundarios (la descripción, la organización de información) (**Cordón-García** *et al*., 2010). La enciclopedia recoge y ordena objetos de conocimiento, organizados en **entradas**. Estos objetos de conocimiento son de naturaleza diversa: personas, acontecimientos, instituciones, conceptos y cualquier elemento nombrable en cualquiera de los campos del conocimiento y la cultura. En la enciclopedia clásica en tomos de papel, las entradas son muy concisas y ordenadas alfabéticamente, dado que ha sido el mecanismo de acceso directo aplicable con éxito a la edición en papel.

Una enciclopedia presenta un importante parecido con un diccionario, no solo en sus manifestaciones impresas, sino también en la versión digital. Esto se debe, por un lado, al acceso alfabético. Por el otro lado, la familiaridad entre diccionario y enciclopedia se debe a que ambos están basados en palabras. El diccionario está orientado a los significados y uso en la lengua de las palabras aisladas (banco) o con especificaciones (banco central, banco de arena); palabras de todo tipo como verbos, adjetivos, adverbios y sustantivos, para una expresión correcta en una lengua. El diccionario acoge, en cierto modo, nombres comunes, y excluye nombres propios o acontecimientos históricos, que son esenciales para la enciclopedia. Lo que podríamos llamar instancias, casos o entidades forman el armazón de las enciclopedias: personajes históricos, lugares, instituciones y creaciones. Un híbrido es el diccionario enciclopédico, en el que se combinan significados (lenguaje) y datos factuales (historia-cultura).

La enciclopedia podría parecer una mera lista de términos, pero también es un esquema de conceptos conectados. Si la enciclopedia es un vocabulario artificial, no puede evitar incorporar parte del aparato de redirecciones y conexiones propias de los vocabularios controlados: Listas de encabezamientos de materias, tesauros, esquemas de conceptos (**Barité**, 2014). En sus gloriosos antepasados en papel también hay destacables esfuerzos de organización sistemática de las ciencias y el conocimiento, que superan el mero ordenamiento alfabético, puesto que este no aporta sentido, y las clasificaciones y esquemas sí.

Tabla 1. Esquema del abordaje de homonimia y polisemia en el lenguaje en sus diferentes escenarios.

Escenario	Explicación	Ejemplo
Diccionarios	Se detallan diferentes acepciones para cada contexto o uso. Cada palabra diferente tiene una entrada propia. Se incluyen, cuando se considera necesario, referencias cruzadas. Suelen incluir listas de sinónimos. A menudo incluyen ejemplos de uso.	*Asiento* m. Asiento, con respaldo o sin él, en que pueden sentarse dos o más personas. m. En los mares, ríos y lagos navegables, bajo que se prolonga en una gran extensión.
Tesauros / Encabezamientos de materias / Esquemas de conceptos	Se establecen denominaciones preferentes y alternativas para un concepto. UP (usado por) y USE Se explica en Nota de Alcance (NA) el significado preciso acordado para un término o concepto.	Automóviles USE coches Animales UP Fauna Fauna USE animales Organización Mundial de la Salud UP OMS OMS USE Organización Mundial de la Salud *Iluminación* NA Incluye tanto la decoración ornamental y las ilustraciones de los manuscritos, como en algunos libros impresos tempranos, si se hizo a mano
Control de autoridades	Se recopilan todas las variantes posibles de una entidad, en varios idiomas y formas de escritura. Se asigna un código único a la entidad. Se tipifican las entidades en personas, organizaciones, lugares y obras	Cartagena (Colombia) Cartagena de Indias Cartagena (España) Museo Nacional Centro de Arte Reina Sofia (España) Leonardo da Vinci, 1452-1519. | Mona Lisa *http://viaf.org/viaf/180757281*
Enciclopedia	Creación de títulos con especificadores Creación de páginas de desambiguación Creación de redirecciones	Cicatriz (novela) América (desambiguación) «Tragicomedia de Calisto y Melibea», Redirección a “La Celestina”

En su versión más usada actualmente (como la *Wikipedia*), la enciclopedia es un conjunto de entradas individuales con un título diferente, accesibles mediante búsqueda en texto libre como sustituto mucho más potente del acceso alfabético. Además, es un conjunto de artículos **densamente conectados** mediante enlaces desde su contenido textual y etiquetas-categorías que agrupan contenidos para la navegación temática por conjuntos de elementos similares. La enciclopedia es, por tanto, también **un mapa del conocimiento**, y por ello comparte aspectos con los sistemas de organización del conocimiento (*KOS*) para establecer conexiones y relaciones significativas. Y, en este sentido, también tiene que lidiar, con un estilo propio, con la **ambigüedad de las palabras para transportar significados** (**Broughton**, 2006).

Los problemas clásicos de homonimia y polisemia en el lenguaje tienen diferente abordaje en cada escenario, los cuales podemos esquematizar así (ver tabla 1) de forma simplificada, teniendo en cuenta que en la enciclopedia se combina la necesidad de identificación unívoca de entidades (personas, organizaciones, lugares) y diferenciación de sentidos (taco, papa).

3. Títulos, redirecciones y desambiguación en *Wikipedia*

Hemos hecho un recorrido superficial por varios instrumentos de organización del conocimiento y terminología que aplican diferentes formas de manejar la naturaleza ambigua del lenguaje natural o conceptos. ¿A qué conceptos puede remitir una palabra inocente como *Once*? A un musical, a un barrio, a un número, a una fundación, a una comida, a una estación, a un disco, a un personaje de *Stranger Things*, etc.

Los diccionarios tienen que enfrentarse a la polisemia, homonimia, homofonía, a través de herramientas aplicadas en la práctica lexicográfica. Las taxonomías y vocabularios controlados usan de forma extensiva las redirecciones entre términos, variantes o usos para orientar a quienes los utilizan en el manejo preciso de cada concepto. También los registros de autoridad para nombres de personas y entidades se construyen con un aparato denso de referencias cruzadas entre nomenclaturas, para manejar la correspondencia entre nombres populares y nombres oficiales, transformaciones organizativas a lo largo del tiempo o relaciones significativas. Sistemas de referencias cruzadas y reenvíos son también parte de los diccionarios bilingües, de aprendizaje de idiomas o de recursos para la traducción.

Las enciclopedias aplican sistemas de referencias cruzadas procedentes del universo de los vocabularios controlados, el control de autoridades y los diccionarios. Aunque la atención mayoritaria para el control de autoridades se la lleva *Wikidata* (**Bianchini**; **Sardo**, 2022), diferentes trabajos también han observado la especificidad de *Wikipedia* como un reto al marco de los vocabularios controlados y el control de autoridades (**Jiménez-Pelayo**,

Las enciclopedias aplican sistemas de referencias cruzadas procedentes del universo de los vocabularios controlados, el control de autoridades y los diccionarios

2009). En este sentido, este trabajo pretende hacer visible el singular y vasto aparato de desambiguación enciclopédico. En general, las obras de referencia se sostienen en el manejo de referencias cruzadas entre conceptos similares y relacionados, para poder sintetizar la variabilidad e imprecisión de las búsquedas en lenguaje natural, tanto en obras basadas en la ordenación alfabética como en las de ordenación sistemática o clasificatoria (**Rodríguez-Yunta**, 2012). Las enciclopedias presentan un contenido muy discursivo, de forma que su núcleo está compuesto de artículos explicativos que requieren lectura y no meramente una consulta puntual. Su contenido no es atómico, en el sentido de las bases de datos, sino que es textual. Los diccionarios tienen un carácter más atómico, basado en variantes de definiciones para cada contexto o caso de uso, en la forma de información textual muy condensada.

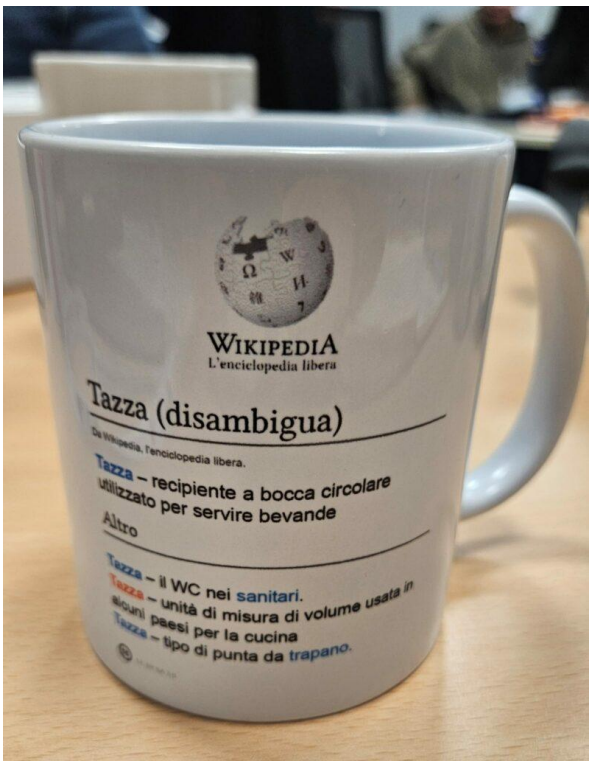


Figura 3. Fotografía de una taza realizada por *Wikimedia Italia*, haciendo honor a la importancia de la desambiguación en *Wikipedia*. *Progetto GLAM Wikipedia e archeologia*, Verona, enero 2025.

El caso de *Wikipedia*, la enciclopedia online que en la actualidad se ha establecido como 'patrón oro' para el concepto de fuente de información enciclopédica generalista (**Reagle**; **Koerner**, 2020), sobre todo por su nivel de uso y ubicuidad digital, presenta una manifestación singular del carácter discursivo de una enciclopedia y del sistema de referencias entre conceptos y términos similares, que se denomina páginas de desambiguación, que es el que vamos a analizar en detalle en este artículo.

Para el lector con una familiaridad solo superficial con *Wikipedia*, conviene describir algunos de sus aspectos definitorios. *Wikipedia* es un campo de estudio

desde muchas perspectivas, abundante y de variadas metodologías. Su éxito durante décadas ha generado un conocimiento generalista, accesible y organizado con muchísimas aplicaciones directas e indirectas (**Tramullas**, 2015).

Hay que resaltar que *Wikipedia* se trata en realidad de un conjunto de enciclopedias escritas en varios idiomas por comunidades de editores diferentes, pero que comparten una serie de principios como el punto de vista neutral, la colaboración horizontal, el uso de fuentes fiables y la licencia de contenido libre. También comparten infraestructuras para la publicación y el software de gestión de contenidos *MediaWiki* que permite a los editores crear, revisar y vigilar los artículos con seguridad y facilidad. Además, los artículos están conectados a una base de conocimiento *Wikidata* que almacena datos factuales estructurados sobre ellos, y que cumple, además, la función crítica de conectar los artículos en diferentes idiomas (**Vrandečić**, 2020).

Wikipedia es un conjunto de enciclopedias escritas en varios idiomas por comunidades de editores diferentes, pero que comparten una serie de principios como el punto de vista neutral, la colaboración horizontal, el uso de fuentes fiables y la licencia de contenido libre

Para acercarnos a la maquinaria que hace funcionar *Wikipedia* tenemos que acercarnos a varios conceptos, siendo el principal el de 'página wiki'. Casi todo el contenido de una *Wikipedia* se materializa en páginas wiki, en las que se van realizando ediciones de texto sucesivas (revisiones) que quedan registradas y accesibles individualmente (historial). Estas páginas son los artículos, escritos usando el lenguaje de marcado ligero *wikitexto*, con el que se formatean resaltados, listas, tablas, referencias, etc. Todas las páginas wiki pueden tener un complemento conocido como “página de discusión”, que permite debatir sobre el contenido y detalles del artículo. Los artículos forman parte de un espacio de nombres por defecto o principal, pero existen otros complementarios como el de categorías, comunidad *Wikipedia* o plantillas. Tanto las categorías, como las plantillas y las páginas internas del portal de editores son también simplemente páginas wiki, con las mismas funciones de serie (**Hernández-Toribio**; **Claes**; **Deltell**, 2020).

Una de las herramientas más útiles para los usuarios avanzados son las plantillas (*A:Plantillas*). Se trata de fragmentos de código que definen la estructura descriptiva básica en artículos con un tipo de contenido concreto (por ejemplo, equipos de baloncesto, óperas, presidentes, etc.), lo que se conoce como fichas o *infoboxes*, y también mensajes de etiquetado y alerta para artículos (Necesita referencias; Artículo en desarrollo; Fallecimiento reciente). Además de las existentes, los editores pueden crear sus propias plantillas.

Desde el punto de vista de los elementos usados en la identificación de los contenidos (puntos de acceso principal, diríamos en terminología clásica de catalogación), en *Wikipedia* encontramos:

- Títulos, para identificar con precisión cada concepto tratado. Existen convenciones para construir los títulos en cada situación: lugares, acontecimientos, plantas, etc.
- Redirecciones de variantes de títulos.
- Páginas de desambiguación, para aclarar diferentes conceptos de título parecido.

Dado que la *Wikipedia* se comporta finalmente como un repertorio alfabético, es el título elegido para cada concepto, hecho o cosa lo que sirve como identificador primario. La página *https://es.wikipedia.org/wiki/Aguamarina* tiene como identificador el título "Aguamarina", y también, aunque invisibilizado, el identificador de página 229922 y el código Q58959 en *Wikidata* que permite acceder a datos sobre esta piedra preciosa desde *https://www.wikidata.org/wiki/Q58959*, así como conectar con artículos equivalentes en 60 idiomas.

Una primera consecuencia de esto son las páginas de redirección entre diferentes formas y variaciones ortográficas de un mismo nombre, usando la función REDIRECT. Las redirecciones son también simples páginas wiki. Por ejemplo, la entrada "Mozart" es una página auxiliar de *Wikipedia* que contiene tan solo el código con la redirección a la forma completa del nombre del compositor (figura 4). Se trata de redireccionamientos unívocos que además a menudo son invisibles para el usuario final, con el código #REDIRECCIÓN [[Wolfgang Amadeus Mozart]]. Aunque no son tan visibles, son muy abundantes, en las estadísticas de *Wikisacan*, se detectan en la *Wikipedia* en italiano 1.813.867 redirecciones.

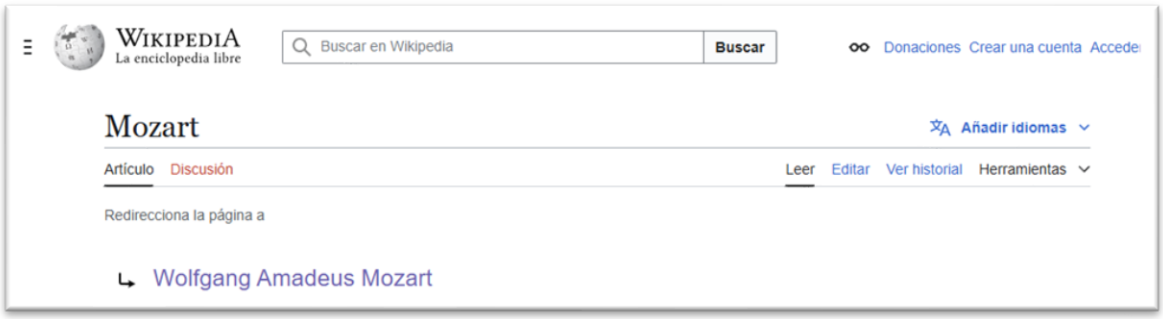


Figura 4. Página de redirección en español para el término Mozart
https://es.wikipedia.org/wiki/Mozart

Pero más allá de estas redirecciones unívocas, las convenciones para títulos (*WP:CT*) establecen un conjunto de pautas para diferenciar artículos con nombre parecido, que busca expresamente "precisión frente a ambigüedad". Así, por ejemplo, para la película *Blade Runner*, el título desnudo corresponde a la película original de 1982, mientras que para otros usos se especifica mediante paréntesis, como: *Blade Runner* (álbum), *Blade Runner* (franquicia) o *Blade Runner* (videojuego de 1997). Sin embargo, hay otros artículos que forman una constelación significativa, como en este caso podría ser: *Blade Runner Trilogy*. *25th Anniversary*, o la novela *Blade Runner*, que corresponde a una edición posterior renombrada del relato *Do Androids Dream of Electric Sheep*? de Philip K. Dick, novela en

la que se basó parcialmente la película, o el artículo del remake *Blade Runner 2049*, o el artículo que detalla las Versiones de *Blade Runner*. En casos como estos se hace necesaria una página de desambiguación (figura 5).

Figura 5. Página de desambiguación en español para el término *Blade Runner* *https://es.wikipedia.org/wiki/Blade_Runner_(desambiguaci%C3%B3n)*

4. Anatomía de las páginas de desambiguación en *Wikipedia*

Una vez explicados los títulos y las redirecciones, podemos entrar en más detalle en las páginas de desambiguación. En las estadísticas de cada enciclopedia, las páginas de desambiguación no cuentan en el total de artículos enciclopédicos de cada *Wikipedia*, sino que se consideran elementos auxiliares, al igual que las páginas de categorías y otros espacios de nombres secundarios.

Recordemos que estas, también, son páginas wiki convencionales; es decir, páginas de texto y enlaces elaboradas conforme a unas pocas convenciones de estructura y formateo en *wikitexto*. Existen unas recomendaciones dentro del manual de estilo de la enciclopedia (*WP:DES*). Por lo general, estas páginas comienzan insertando una plantilla específica, en español {{Desambiguación}} y existe una versión de esta plantilla para 208 idiomas: en inglés se denomina {{Disambiguation}}, en francés {{Homonymie}} y en alemán {{Begriffsklärung}}.

Las plantillas son “la navaja suiza de la *wiki*” (**Saorín**, 2012, p. 78). En el caso que nos interesa cumple dos funciones: la principal es que permite insertar el mismo mensaje normalizado en todas las páginas, mediante lo que en el software *MediaWiki* se llama transclusión: el copiado automático del contenido y diseño de la plantilla para generar un aviso al comienzo o al final de la página, que por lo general se identifica con el símbolo icono de desambiguación.

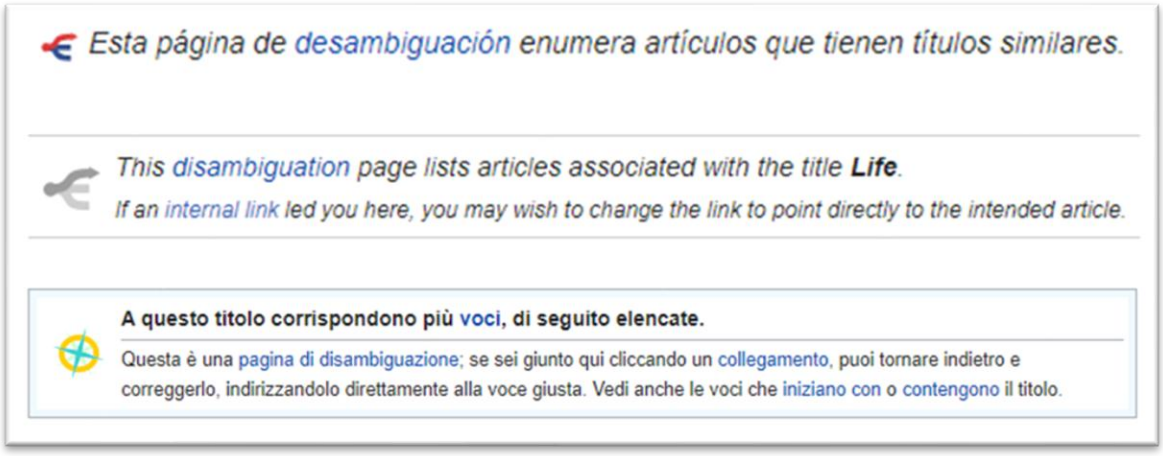


Figura 6. Tres ejemplos del aspecto del aviso producido en las páginas desde la plantilla de desambiguación en español, inglés e italiano.

Tanto el símbolo como el texto informativo varía en extensión y detalles según la versión de la enciclopedia: explica que se trata de un listado o enumeración de artículos que "tienen títulos similares" o "associtated with the title …" o "différents sujets et articles partageant un même nom..." Es frecuente que se incluya el consejo de corregir enlaces si desde un artículo se ha enlazado, por descuido, a una página de desambiguación en lugar de al destino preciso. La otra función que cumple la plantilla de desambiguación es que se encarga de categorizar la página como "Página de desambiguación", lo cual permite identificarla como una página "no artículo" en la *Wikipedia*. Esta forma de tipificación original se establece también en *Wikidata*, instanciando el elemento correspondiente a cada página de desambiguación con la declaración P31 > Q4167410.

Por otro lado, aunque no se tratará aquí, existen versiones de esta plantilla para su uso en otros proyectos como *Wikimedia Commons* para las categorías, *Wikidata*, *Wikinoticias*, *Wikiquote*, *Wikisource*, etc.

Las páginas de desambiguación son discursivas, se materializan en un texto formateado a criterio del editor. No son una base de datos y, por lo tanto, no podríamos construir desde ellas un mapa de conceptos similares ni otros procesos automáticos.

Suelen comenzar con una breve línea que explica que el término X puede referirse a diferentes conceptos, hechos o cosas, y a continuación se presenta un simple listado de viñetas. A continuación de cada enlace se suele añadir una concisa explicación del sentido del artículo, que es lo que podemos considerar la desambiguación en sí misma.

Las páginas de desambiguación son discursivas, se materializan en un texto formateado a criterio del editor. No son una base de datos y, por lo tanto, no podríamos construir desde ellas un mapa de conceptos similares ni otros procesos automáticos.

Babel (desambiguación)

29 idiomas

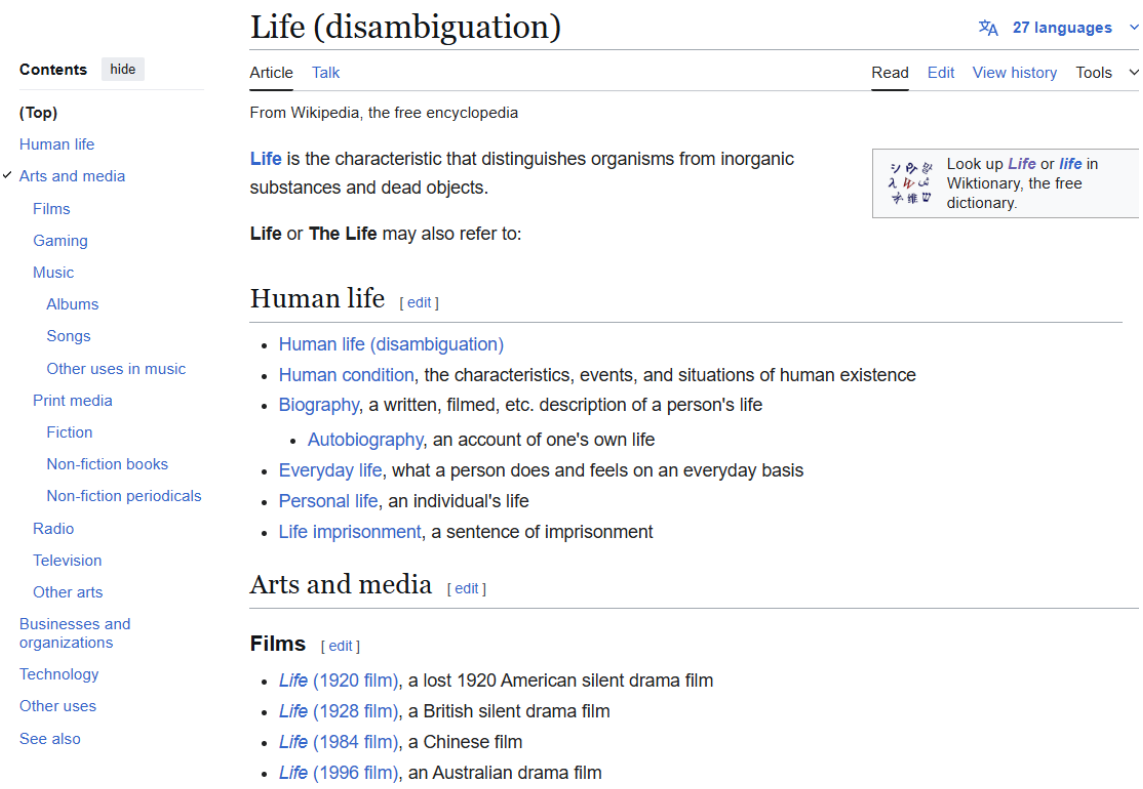
Artículo Discusión Leer Editar Ver historial Herramientas

Esta página de desambiguación enumera artículos que tienen títulos similares.

Término proveniente del hebreo **Bābel**. Babel puede referirse a:

- **Babilonia**;
- la **Torre de Babel**;
- **Babel**, la ciudad de Babilonia desde una perspectiva bíblica;
- ***Babelfish***, aplicación de Internet;
- ***La biblioteca de Babel***, cuento de Jorge Luis Borges;
- ***Babel***, película de Alejandro González Iñárritu;
- ***Babel***, revista literaria argentino-chilena;
- al álbum ***Babel***, del grupo español Azul y Negro;
- al álbum ***Babel***, el primero de la banda de rock colombiana The Mills;
- al álbum ***Babel***, de la banda inglesa de folk *rock* Mumford & Sons;
- **Pez de Babel**, una especie de pez de ficción que aparece en *Guía del autoestopista galáctico* de Douglas Adams;
- **Wikipedia:Babel**, en esta página se explica cómo informar a los demás de las lenguas que habla cada usuario. Esta iniciativa facilita la comunicación dentro de una comunidad plurilingüe.

Figura 7a. Ejemplo de página de desambiguación con listado simple y artículos en rojo (Babel, en español)



Life (disambiguation)

27 languages

Contents hide

(Top)
Human life
Arts and media
Films
Gaming
Music
Albums
Songs
Other uses in music
Print media
Fiction
Non-fiction books
Non-fiction periodicals
Radio
Television
Other arts
Businesses and organizations
Technology
Other uses
See also

Article Talk Read Edit View history Tools

From Wikipedia, the free encyclopedia

Life is the characteristic that distinguishes organisms from inorganic substances and dead objects.

Look up *Life* or *life* in Wiktionary, the free dictionary.

Life or **The Life** may also refer to:

Human life [edit]

- Human life (disambiguation)
- Human condition, the characteristics, events, and situations of human existence
- Biography, a written, filmed, etc. description of a person's life
  - Autobiography, an account of one's own life
- Everyday life, what a person does and feels on an everyday basis
- Personal life, an individual's life
- Life imprisonment, a sentence of imprisonment

Arts and media [edit]

Films [edit]

- *Life* (1920 film), a lost 1920 American silent drama film
- *Life* (1928 film), a British silent drama film
- *Life* (1984 film), a Chinese film
- *Life* (1996 film), an Australian drama film

Figura 7b. Ejemplo de página de desambiguación con listado agrupado por temas (*Life*, en inglés)

Cuando una página contiene muchas remisiones, los artículos enumerados pueden venir estructurados en epígrafes según temáticas, en cuyo caso la página presenta además un índice de contenido (TOC: *table of content*). Conforme a la mecánica wiki, los artículos a los que se remiten pueden no existir todavía, y estar solo indicados, apareciendo en rojo, actuando así la página de desambiguación como un recordatorio de nuevos artículos pendientes de elaborar.

Conviene recordar que la desambiguación, al igual que los artículos, es particular para cada idioma (**Jemielniak**; **Wilamowski**, 2017). La forma y los elementos que participan en una página de desambiguación en la edición alemana son diferentes de su homóloga en italiano o francés. Pero, aparte del formato, lo más relevante es que son las prácticas editoriales de quienes editan *Wikipedia* las que determinan su cobertura, precisión, redacción y otros detalles de lo que podríamos llamar un "nanoartículo de orientación", puesto que no se trata de una referencia cruzada formalizada, sino de un texto explicativo en el que se pueden recoger observaciones y puntualizaciones. Estas páginas son, en cierto modo, un reflejo del cuidado editorial para señalar a los conceptos adecuados y facilitar el descubrimiento del artículo adecuado a la necesidad de información de cada lector.

Además de en las propias páginas, quienes editan también pueden realizar intervenciones en los artículos afectados por ellas . En los artículos de *Wikipedia* que pueden suscitar dudas, debidas a la homonimia o cercanía de dos expresiones o títulos de artículos, se recomienda añadir la mención a la página de desambiguación en la que se aclaran las diferentes acepciones con las que puede confundirse la entrada. Esto se realiza mediante una plantilla, que en español es la de {{Otros usos}}, que puede apuntar a una o varias páginas de desambiguación, según el criterio del editor.

A propósito de las páginas de desambiguación, son las prácticas editoriales de quienes editan *Wikipedia* las que determinan su cobertura, precisión, redacción y otros detalles de lo que podríamos llamar un "nanoartículo de orientación", puesto que no se trata de una referencia cruzada formalizada, sino de un texto explicativo en el que se pueden recoger observaciones y puntualizaciones

Un aspecto sintomático de la dificultad de la edición colaborativa descentralizada en *Wikipedia* es el uso erróneo de enlaces desde artículos a páginas de desambiguación. Los *wikienlaces* en los artículos no pueden contener ambigüedad, quien edita ha de seleccionar como destino el concepto adecuado y nunca remitir a una página de desambiguación. Sin embargo, este descuido sucede a menudo, y por ello, en uno de los "Wiki Fixup projects", un bot "patrulla" artículos para detectar este tipo de enlaces imprecisos y arreglarlos.

Figura 8. Ejemplo de la inserción de {{Otros usos}} en el artículo de América, sobre el continente americano, mediante el código wiki {{Otros usos|América (desambiguación)|otro=Americano (desambiguación)}}

Tabla 2. Total de artículos y de páginas de desambiguación en las *Wikipedias* con más artículos.

Wiki	Nº páginas desambiguación	Nº total artículos	Ratio
English	340.471	6.984.942	5%
Alemán	333.886	3.008.591	11%
Ruso	190.145	2.040.810	9%
Francés	155.066	2.679.173	6%
Italiano	130.787	1.914.645	7%
Neerlandés	94.181	2.185.026	4%
Polaco	92.140	1.655.360	6%
Japonés	82.984	1.459.768	6%
Ukranio	79.398	1.374.656	6%
Sueco	79.234	2.608.848	3%
Coreano	64.630	702.051	9%
Español	59.940	2.027.494	3%
Chino	57.363	1.472.897	4%
Serbio	43.013	706.074	6%
Árabe	33.442	1.259.703	3%
Húngaro	31.843	556.389	6%
Portugués	31.492	1.145.896	3%
Serbocroata	30.579	460.612	7%
Checo	26.742	567.285	5%
Farsi	24.178	1.036.561	2%
Eslovaco	23.855	253.357	9%
Finés	21.355	592.717	4%
Catalán	19.896	772.988	3%

5. ¿Cuánto y qué se desambigua en *Wikipedia* en cada idioma?

Hasta aquí la explicación del procedimiento para desambiguar. A partir de aquí podemos acercarnos a datos reales sobre su verdadero uso en *Wikipedia*. Para los datos de páginas existentes es sencillo de obtener mediante consultas SPARQL a *Wikidat*a. Para otros datos, como la participación de editores y artículos enlazados hay que desplegar procedimientos más farragosos a través de scripts Python de consulta a la API de *Wikipedia* o técnicas de scrapping, según criterios de eficiencia.

Entre todas las *Wikipedias* encontramos 1.502.265 páginas de desambiguación, pero hay que tener en cuenta que los datos son específicos de cada idioma. Un primer análisis exploratorio de los resultados señala diferencias significativas entre ediciones en la ratio entre páginas de desambiguación y total de artículos. Hemos obtenido una ratio entre el número de páginas desambiguación y el total de artículos en las mayores enciclopedias, siendo el dato promedio general poco superior al 5%. Los datos a 22 de abril de 2025 se muestran en la tabla 2.

Como es de esperar, hay una correlación entre la cantidad total de artículos y la cantidad de páginas de desambiguación. No todas las *Wikipedias* se comportan igual, especialmente la alemana, con la ratio más alta (11%) y que se aleja de la línea de tendencia (Gráfico 1).

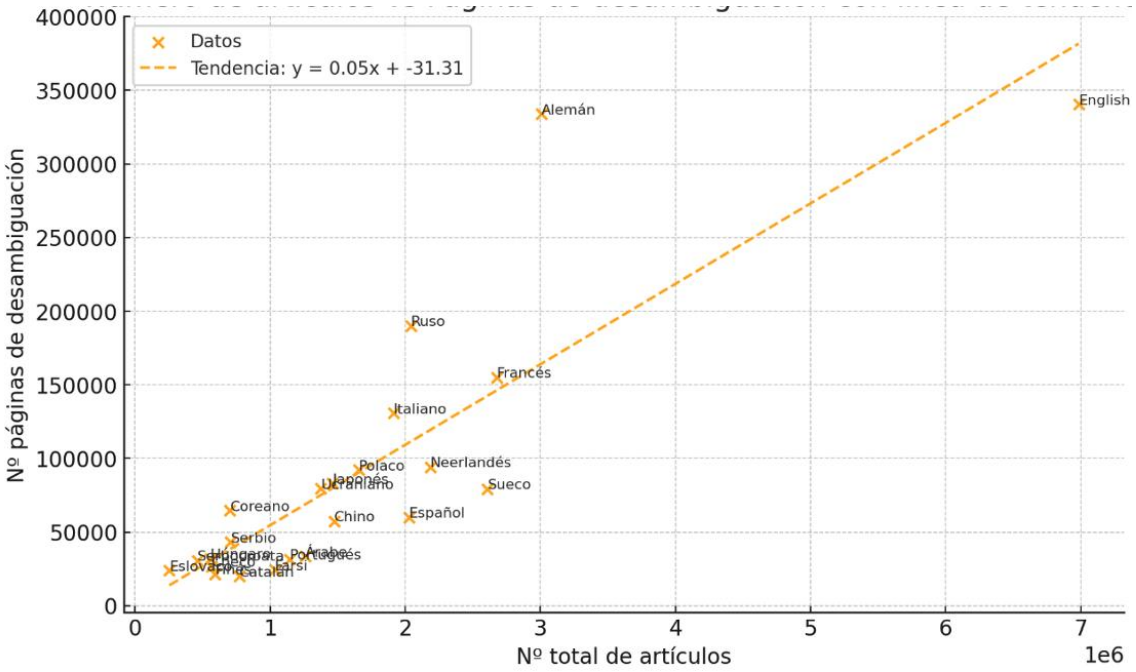


Gráfico 1. Diagrama de dispersión de número de páginas de desambiguación *vs* número total de artículos por idioma.

Para acercarnos a una visión de la desambiguación como tarea realizada por editores/as de *Wikipedia*, hemos indagado también en los datos de editores que crean y modifican estas páginas. A través de la API de *MediaWiki* hemos procesado los datos de editores únicos de *Wikipedia* en español.

Los datos de editores humanos muestran que la gran mayoría de los editores (el 96%; 170,443 usuarios) realiza un número muy reducido de contribuciones, mientras que un grupo extremadamente pequeño concentra una cantidad elevada de ediciones (3%; 6,013 editores realizaron entre 3 y 10 ediciones). Muy pocos son los editores altamente activos (más de 300 ediciones), pero seguramente desempeñan un papel crucial en la sostenibilidad y continuidad del contenido, asumiendo tareas de mantenimiento y cohesión. Se podría enriquecer esta observación con otras características de estos editores, especialmente si tienen roles de 'bibliotecario' u otros permisos especiales de supervisión. Mención aparte merecen los *bots*.

Otra exploración complementaria es indagar sobre el total de artículos afectados por desambiguación, teniendo en cuenta que una página de desambiguación remite a X artículos. Al realizar los cálculos, es llamativo ver que las páginas afectadas por desambiguación son un altísimo porcentaje del total de cada *Wikipedia*. La obtención de estos datos puede hacerse mediante técnicas de scrapping o interrogando a la API de *Wikipedia*. La segunda opción es un proceso más lento, pero los datos son más rigurosos y es la que hemos usado para obtener datos de *Wikipedia* en español. La exploración se puede extender con otros detalles, como si se han insertado puntualmente en cada página la plantillas o marcas de desambiguación, o el número de “enlaces rojos” a artículos no existentes desde páginas de desambiguación.
Para la *Wikipedia* en castellano, hemos analizado los artículos a los que se enlaza desde las páginas de desambiguación. Se detectaron más de seiscientos mil enlaces, de los que el 89% son a artículos (hay un pequeño porcentaje de enlaces a artículos de anexo, páginas de ayuda y de políticas internas de edición en *Wikipedia*). Depurando los duplicados nos encontramos con una contundente cantidad de 405.495 artículos propensos a la ambigüedad, un 20% de todos los artículos de la enciclopedia. También encontramos que desde las páginas de desambiguación se reclama la creación de más de ochenta mil artículos (conocidos como enlaces en rojo).

La página de desambiguación promedio en español incluye 9,09 referencias a artículos. Algunos artículos son desambiguados desde varias páginas, pero el 74% desde solo una. En cada página, el promedio es de 1,34 enlaces rojos que apuntan a artículos que necesitan ser creados (tabla 3).

Tabla 3. Datos de *Wikipedia* en español sobre artículos y desambiguación.

Artículos desambiguados (total)	**Artículos desambiguados (únicos)**	**Ratio ambiguos únicos sobre total**	**Enlaces rojos**
544.676	405.495	20%	80.489

Sería interesante preguntarse a qué categorías pertenecen estos más de cuatrocientos mil artículos. En una primera aproximación encontramos que más de un tercio corresponden a personas. Del mismo modo, una comparación de la equivalencia entre idiomas en las páginas de desambiguación, nos daría otra intere-

sante visión sobre diversidad cultural y los llamados “content imbalances” (**Miquel-Ribé**; **Laniado**, 2021). En una exploración preliminar, hemos visto que el 23% de las páginas de desambiguación de la *Viquipèdia* en catalán no existen en ningún otro idioma.

6. Reflexiones finales

En una enciclopedia podemos considerar la elaboración y mejora del contenido de los artículos como el trabajo primario, y considerar como trabajos auxiliares a otros como la anotación con plantillas, la categorización o la creación de páginas de desambiguación. En este sentido, estos trabajos secundarios nos pueden estar hablando de capas de calidad y atención que podrían pasar desapercibidas a simple vista, pero que desde el punto de vista de los trabajos y las herramientas de organización del conocimiento cobran bastante importancia. Entonces, ¿podríamos sugerir que los alemanes son mejores editores de *Wikipedia* porque categorizan, anotan y desambiguan más?

Por otra parte, dado que la desambiguación está muy relacionada con la forma de los nombres, podría considerarse si ciertos idiomas presentan más propensión a la ambigüedad que otros, aunque no parece una afirmación fácilmente sostenible. Para explicar las diferencias entre idiomas, nos inclinamos más al resultado de variaciones culturales sin explicación clara, y, seguramente, a la tendencia marcada por algunos editores muy activos, con larga experiencia.

El seguimiento de las pautas de estilo y procedimiento en ciertas tareas secundarias puede variar bastante entre una comunidad de editores y otra, lo cual recuerda el importante papel que juegan en las *Wikipedias* los bots que patrullan y detectan errores de detalle, poco visibles al ojo humano, y que rematan tareas muy minuciosas.

Finalmente, planteamos también otro aspecto derivado de la madurez de *Wikidata* como base de conocimiento enciclopédico. A día de hoy, las páginas de desambiguación están registradas como tales en *Wikidata*, pero no así la información que vehiculan. En una exploración inicial encontramos que apenas están esbozados algunas desambiguaciones, usando las propiedades P1889 (diferente de) y P460 (Se dice que es lo mismo que), pero sin capturar la especificidad de cada desambiguación según el idioma en el que tiene lugar. Parece que sería propicio para una formalización más rigurosa. Se adivinan dificultades en la modelización, ya que la desambiguación es diferente en cada idioma, al estar relacionada con la existencia de títulos similares, como puede verse en la página en español para "Once" y las correspondientes para "Once" en inglés, alemán o francés. También sería interesante, desde la perspectiva clásica del reconocimiento de entidades nombradas (NER) usar la tipificación registrada en *Wikidata* para clasificar, entre los artículos enlazados desde páginas de desambiguación, cuáles son de personas, organizaciones o lugares. Se trata de aspectos que requieren “further research”, si alguien lo encontrase interesante.

7. Referencias

A:Plantillas, Ayuda:Plantillas. *https://es.wikipedia.org/wiki/Ayuda:Plantillas*

Barité-Roqueta, Mario-Guido (2014). El control de vocabulario en la era digital: revisión conceptual. *Scire*. 20:1 (en.-jun. 2014) 99-108.
https://doi.org/10.54886/scire.v20i1.4196

Bianchini, Carlo; **Sardo, Lucia** (2022). Wikidata: a new perspective towards universal bibliographic control. *JLIS*, 13(1).
https://doi.org/10.4403/jlis.it-12725

Broughton, Vanda (2006). *Essential thesaurus construction*. London: Facet Publishing. ISBN: 978 1 856045650

Cordón-García, José-Antonio; **Alonso-Arévalo, Julio**; **López-Lucas, Jesús**; **Gómez-Díaz, Raquel** (2010). Las enciclopedias: el caso particular de la Wikipedia. En: *Las nuevas fuentes de información: información y búsqueda documental en el contexto de la web 2.0*. Madrid: Pirámide.

Hernández-Toribio, María-Isabel; **Claes, Florencia**; **Deltell, Luis** (2020). Español/castellano en *Wikipedia*: voces y foro de discusión. *Doxa Comunicación. Revista Interdisciplinar de Estudios de Comunicación y Ciencias Sociales*, *31*, 283-302.
https://doi.org/10.31921/doxacom.n31a14

Jemielniak, Dariusz; **Wilamowski, Maciej** (2017). Cultural diversity of quality of information on Wikipedias. *ASIS&T*, 68(10), pp. 2460–2470.
https://doi.org/10.1002/asi.23901

Joudrey, Daniel N.; **Taylor, Arlene G.** (2018). *The organization of information* (4th ed.). Santa Barbara, California: Libraries Unlimited. ISBN: 978 1 598848588

Miquel-Ribé, Marc; **Laniado, David** (2021). The Wikipedia Diversity Observatory: helping communities to bridge content gaps through interactive interfaces. *Journal of Internet Services and Applications*, vol. 12, 10.
https://doi.org/10.1186/s13174-021-00141-y

Reagle, Joseph; **Koerner, Jackie** (eds.) (2020). *Wikipedia @ 20: Stories of an Incomplete Revolution*. MIT Press.

Rodríguez-Yunta, Luis (2012). Control de autoridades, una herramienta desaprovechada en los sistemas de recuperación. *Anuario ThinkEPI*, v. 6, pp. 240-243.
https://thinkepi.scimagoepi.com/index.php/ThinkEPI/article/view/30426/16004

Saorín-Pérez, Tomás (2012). *Wikipedia de la A a la W*. Colección EPI-UOC. Barcelona: UOC.

Tramullas, Jesús (2015). *Wikipedia* como objeto de investigación. *Anuario ThinkEPI,* v. 9, pp. 223-226.
https://doi.org/10.3145/thinkepi.2015.50

Vrandečić, Denny (2020). Collaborating on the sum of all knowledge across languages". In: Reagle, Joseph; Koerner, Jackie (eds.). *Wikipedia @ 20: Stories of an Incomplete Revolution*. MIT Press, 2020.
https://wikipedia20.mitpress.mit.edu/pub/svu3uy1z/release/2

WP:CT, Wikipedia:Convenciones de títulos. *https://es.wikipedia.org/wiki/WP:CT*

WP:DES, Wikipedia:Manual de estilo/Páginas de desambiguación. *https://es.wikipedia.org/wiki/WP:DES*