TENDÊNCIAS ATUAIS E PERSPETIVAS FUTURAS EM ORGANIZAÇÃO DO CONHECIMENTO

ATAS DO III CONGRESSO ISKO ESPANHA-PORTUGAL
XIII CONGRESSO ISKO ESPANHA

Universidade de Coimbra, 23 e 24 de novembro de 2017

Com a coordenação de

Maria da Graça Simões, Maria Manuel Borges

TÍTULO
Tendências Atuais e Perspetivas Futuras em Organização do Conhecimento: atas do III Congresso ISKO Espanha e Portugal - XIII Congresso ISKO Espanha

COORDENADORES
Maria da Graça Simões
Maria Manuel Borges

EDIÇÃO
Universidade de Coimbra. Centro de Estudos Interdisciplinares do Século XX - CEIS20

ISBN
978-972-8627-75-1

ACESSO
https://purl.org/sci/atas/isko2017

OBRA PUBLICADA COM O APOIO DE

PROJETO UID/HIS/00460/2013

UTILIZACIÓN DE CATEGORÍAS DE WIKIPEDIA EN PROCESOS DE ORGANIZACIÓN DE INFORMACIÓN: HACIA UNA REVISIÓN CUALITATIVA.

Jesús Tramullas[1], Ana I. Sánchez-Casabón[2], Piedad Garrido-Picazo[3]

[1]Depto. Ciencias de la Documentación, Univ. de Zaragoza, 0000-0002-5374-9993, tramullas@unizar.es
[2]Depto. Ciencias de la Documentación, Univ. de Zaragoza, 0000-0002-0908-1615, asanchez@unizar.es
[3]Depto. Informática e Ingeniería de Sistemas, Univ. de Zaragoza, 0000-0002-1750-7225, piedad@unizar.es

RESUMEN Este trabajo revisa el estudio y la utilización del sistema de categorías de Wikipedia en la investigación científica, adoptando para ello una metodología de revisión sistemática de literatura, pero atendiendo a la revisión cualitativa del contenido de un conjunto de trabajos escogidos. Se identifican varios tipos de trabajos, atendiendo al estudio intrínseco del sistema de categorías, o a su uso como herramienta para el análisis de otros corpus documentales diferentes a Wikipedia. Se concluye que el sistema de categorías ofrece un esquema de clasificación válidos para abordar estudios sobre la organización del conocimiento en múltiples contextos.

PALABRAS CLAVE *categorías, Wikipedia, organización del conocimiento, ontologías, clasificación.*

ABSTRACT This paper reviews the study and use of the Wikipedia category system in the scientific research, adopting a systematic literature review approach, but considering the qualitative review of the content of a set of selected papers. Several types of work are identified, depending on the intrinsic study of the category system, or its use as a tool for the analysis of other documentary corpus other than Wikipedia. We conclude that the system of categories offers a valid classification scheme for different approaches and studies on the organization of knowledge in multiple contexts.

KEYWORDS *categories, Wikipedia, knowledge organization, ontologies, classification*

INTRODUCIÓN

Wikipedia es el principal recurso de información enciclopédica disponible a escala mundial. Ofrece más de 32 millones de artículos, y es consultada diariamente por millones de usuarios. Puede considerarse en la actualidad como la mayor base de conocimiento, que es organizado y etiquetado conforme a unas instrucciones y parámetros básicos. Si se atiende a datos estructurados según criterios semánticos, entonces es el proyecto Wikidata (Vrandecic, 2013) la mayor base de conocimiento existente. La revisión de la estructura de los artículos en Wikipedia, así como las herramientas de organización y exploración de la enciclopedia, permiten identificar que uno de sus elementos fundamentales, inherentes al contenido, son las categorías (figura 1). El conjunto de las categorías, que

está compuesto por un vasto conjunto de términos, y que se encuentra en constante evolución, es utilizado por los editores para enmarcar los contenidos dentro de una estructura de organización del conocimiento (Capocci, Rao & Caldarelli, 2008). Las categorías fueron introducidas en Wikipedia en 2003 y las páginas de categorías y subcategorías en 2004. Ambos elementos son definidos, mantenidos y actualizados por la comunidad de editores de manera colaborativa (Thornton y MacDonald, 2012). Se trata de un sistema que combina una organización jerárquica con relaciones entre diferentes categorías, de forma que se crean polijerarquías y asociaciones. A todos los efectos, el sistema de categorías es un sistema de organización del conocimiento, y permiten agrupar los artículos en conjuntos y subconjuntos conceptuales o temáticos.

El estudio científico de la dinámicas de Wikipedia ha dado como resultado la publicación de trabajos sobre los procesos de edición colaborativos, las pautas de comportamiento de las comunidades de usuarios, el fenómeno del vandalismo, etc. (Mesgari et alii, 2015). El desarrollo de los estudios sobre el universo de Wikipedia ha permitido elaborar y disponer de varias revisiones sistemáticas de literatura, que atienden a diferentes enfoques (Tramullas, 2015). Dentro del conjunto de trabajos sobre Wikipedia, que muestran diferentes orientaciones, objetivos, métodos y resultados, también es posible identificar en la bibliografía especializada un buen número de estudios que están usando el corpus textual de Wikipedia, y en particular las categorías, como componente para diferentes estudios relacionados con la organización y la recuperación de información, el etiquetado social, la clasificación de documentos, o el etiquetado semántico y las ontologías.

Figura 1. Categorías en el artículo de Wikipedia "Alien: el octavo pasajero" (modo Edición).

En consonancia con todos estos aspectos, este trabajo tiene como objetivo principal identificar los usos y las aplicaciones que los investigadores están haciendo del sistema de categorías de Wikipedia, y cómo esta utilización se está reflejando en la evolución del corpus de literatura científica disponible. En segundo lugar, pretende revisar la forma en que un sistema de organización del conocimiento, elaborado colaborativamente, se está usando como recurso de investigación en diferentes aproximaciones al tratamiento y organización de la información, lo que refrendaría su validez como sistema de clasificación. Finalmente, debe indicarse que este trabajo no entra a valorar la estructura, la evolución ni la calidad del sistema de categorías de Wikipedia.

METODOLOGÍA

La metodología utilizada para la realización del trabajo ha sido la revisión sistemática de bibliografía. En este caso se ha optado por un estudio cualitativo, seleccionando trabajos específicos a revisar, antes que por un estudio cuantitativo descriptivo o inferencial de tipo bibliométrico. La recopilación de los datos bibliográficos se ha llevado a cabo a través de la consulta de las referencias sobre el tema objeto

de estudio disponibles en *Web of Science* y *Scopus*. El método de trabajo ha sido adoptado del propuesto por Okoli y Schabram (2009) para el estudio de los trabajos de investigación sobre Wikipedia, y que ya ha sido aplicado previamente por los autores de este trabajo (Tramullas, Garrido-Picazo y Sánchez-Casabón, 2016).

Durante la primera fase se seleccionaron de fuentes y la consulta a realizar. La consultas sobre *Scopus* y *Web of Science* se llevaron a cabo entre febrero y marzo de 2017, usando la expresión "Wikipedia AND categories", y limitando la búsqueda a trabajos publicados entre 2002 y 2016. Se obtuvieron 666 resultados en Scopus y 311 en Web of Science. En ambos casos, los primeros trabajos identificados se publicaron en 2006. No se consultaron, por los objetivos y límites establecidos para este trabajo, ni las bibliotecas digitales de ACM ni IEEE, por la repetición de contenidos, ni *Google Scholar*, por la imposibilidad de limitar las búsquedas a posiciones específicas de los documentos o de sus referencias bibliográficas.

En una segunda fase, una vez obtenidos los datos en bruto de las referencias, se ha procedido a su procesamiento. En primer lugar se han fusionado ambos conjuntos, para proceder a la identificación y eliminación de los documentos duplicados. Esta tarea ha reducido el número de trabajos a 669. Posteriormente, se ha procedido a realizar una selección cualitativa de los trabajos, atendiendo a la identificación temática del contenido de los diferentes estudios, mediante la revisión de los títulos, resúmenes y palabras clave asignados a los mismos. El criterio de selección establecido fue el uso o estudio de categorías como elemento importante del trabajo que se revisase en cada caso. Cada trabajo fue revisado por los tres autores de manera independiente. En caso de discrepancia se revisaba el trabajo para decidir su inclusión o rechazo por mayoría. Este tipo de selección no permite valorar la calidad, pero este criterio no ha sido tenido en cuenta, ya que este trabajo no tiene como objetivos ni elaborar una bibliografía seleccionada ni evaluar la calidad de los trabajos o su impacto. El filtrado realizado ha permitido eliminar del conjunto aquellos artículos o comunicaciones cuyo uso o referencia a las categorías de Wikipedia no estaba directamente relacionado con los objetivos planteados en este trabajo. Finalmente, el número de trabajos seleccionados ha ascendido a 518.

Los estudios y trabajos seleccionados han sido revisados para identificar en los mismos la utilización que se ha hecho del sistema de categorías de Wikipedia. Se han identificado el contexto de aplicación, el uso de las categorías, el método utilizado y los resultados obtenidos, con la finalidad de delinear las líneas de investigación que han usado las categorías de Wikipedia como parte integrante de las mismas. Finalmente, se ha procedido a la descripción de los resultados obtenidos y a la elaboración de la síntesis y las propuestas de desarrollo.

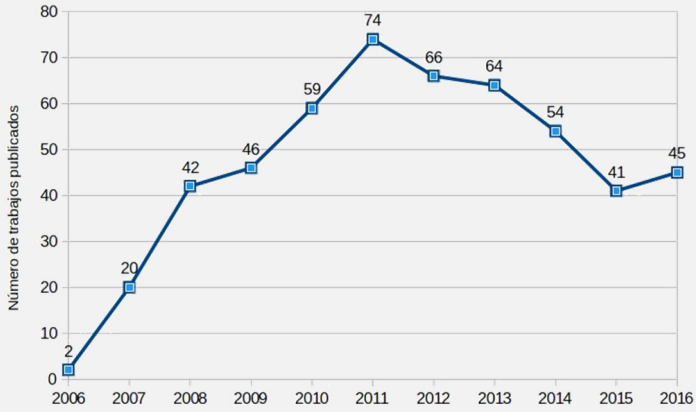

Figura 2. Evolución del número de publicaciones por año (*elaboración propia*).

Todos los datos obtenidos durante el proceso han sido liberados como *Open Data*, y las referencias bibliográficas obtenidas y revisadas se han publicado en grupos específicos de acceso público en *Mendeley*, *Zotero y CiteUlike* (tabla 1).

Tabla 1. Datos en acceso abierto

Mendeley	https://www.mendeley.com/community/research-on-wikipedia-categories/documents/
Zotero	https://www.zotero.org/groups/1543457/research_on_wikipedia_categories
CiteUlike	http://www.citeulike.org/groupdefault/20700

RESULTADOS

Los resultados obtenidos de la revisión cualitativa que se ha llevado a cabo demuestran la gran variedad de enfoque, usos y aplicaciones que los investigadores hacen del sistema de categorías de Wikipedia. Esta riqueza, a su vez, implica un límite para el análisis cualitativo que se propone: la combinación de técnicas, enfoques y aplicaciones existente en los trabajo de investigación imposibilita poder establecer divisiones precisas entre tipos de trabajos. Si bien pueden encontrar trabajos que puede adscribirse claramente a un tema (por ejemplo, generación de ontologías), se encuentran acto seguido trabajos que generan ontologías y las combinan con otras técnicas para poder aplicarlas a procesos de recuperación de información, que, a su vez, puedan darse en un contexto genérico, o estar circunscritos a dominios específicos. En consecuencia, un tratamiento cualitativo como el que aborda este trabajo debe limitarse a delinear temas y líneas de investigación claramente identificados. Para poder investigar las relaciones entre los temas y las posibles divisiones o tipos entre ellos sería necesario aplicar técnicas de procesamiento de información basadas en análisis numéricos del corpus bibliográfico, como recientemente han desarrollado Smiraglia y Cai (2017).

Atendiendo a la limitación indicada, la revisión cualitativa permite, en primer lugar, establecer una primera división. En primer lugar, deben señalarse aquellos estudios que analizan el sistema de categorías dentro del propio contexto de Wikipedia (cubriendo aspectos como la organización y el ciclo de vida de los contenidos, su estructura, la utilización por la comunidad de usuarios, o la propia evolución y mejora del sistema de categorías…). En segundo, aquellos trabajos que utilizan las categorías de Wikipedia en el marco de estudios sobre diferentes aspectos del procesamiento y tratamiento de la información, generalmente sobre corpus documentales ajenos a Wikipedia (formados por colecciones de documentos de diferentes tipos y/o temáticas, páginas web, mensajes en redes sociales…). Alrededor del 90% de los estudios revisados corresponden a este segundo grupo. Hay que destacar la presencia de trabajos que usan corpus creados ex-profeso que, a su vez han sido generados o extraídos de la propia Wikipedia. Todos los trabajos revisados podrían encuadrase en alguna de las cuatro categorías de investigación sobre Wikipedia enumeradas por Nielsen (2017), aunque este investigador sitúa el tema tratado en este trabajo dentro de la categoría genérica de “usos de Wikipedia”.

Dentro de este numeroso grupo de trabajos se pueden establecer varias grandes subdivisiones, aunque, como se ha indicado previamente, hay trabajos que pueden atribuirse a dos o más de ellas. Los grupos que se proponen corresponden a:

1) Recuperación de información: se trata de aquellos trabajos en los cuales las categorías se han usado en diferentes procesos y/o técnicas de recuperación de información, tanto

en lo que concierne a la formulación de expresiones de búsqueda, como su refinamiento y mejora, o al filtrado de resultados. Deben destacarse su uso en procesos de evaluación de resultados. También se han usado en procesos de recomendación.

2) Procesamiento de entidades: especial interés, también por su número, despiertan los trabajos que buscan identificar entidades (*named entities*) en los documentos. La aplicación de estos estudios es muy amplia, ya que sirven para identificar relaciones semánticas entre términos, resolver problemas de desambiguación, o integrar clasificaciones y taxonomías. Pueden aparecer relacionados con investigaciones sobre procesamiento de lenguaje natural e incluso la elaboración de diccionarios en diferentes lenguas o multilingües.
3) Indización y clasificación de corpus documentales: las categorías, o subconjuntos específicos de ellas, han sido usadas como herramienta para proceder a la indización y clasificación de conjuntos de documentos, normalmente dentro de contextos o dominios específicos. Un tipo especial de estos trabajos lo forman aquellos que usan las categorías para etiquetar documentos textuales, en el marco de procesos de indización automática. Otro subconjunto a destacar es el formado por estudios que abordan el etiquetado automático de video o fotografía. Quizá cabe incluir aquí los trabajos que elaboran corpus especializados de forma automática, usando a su vez las categorías de Wikipedia.
4) Creación y uso de taxonomías: se trata de uno de los usos más clásicos. Las categorías de Wikipedia son extraídas de su contexto para formar esquemas nuevos, que puedan ser aplicados en dominios específicos. Pueden ser completadas con el uso combinado de otras taxonomías. En relación con esta aproximación se han identificado algunos trabajos que proponen la creación de clasificaciones de estructura clásica, como clasificaciones jerárquicas o tesauros. En numerosas ocasiones las taxonomías creadas se integran en procesos de indización y clasificación de corpus, como se recoge en el punto 3 de esta enumeración.
5) Creación y uso de ontologías: el segundo de los usos clásicos. Alrededor de un 15% de los trabajos revisados se ocupan de la creación y uso de taxonomías y ontologías desde el sistema de categorías de Wikipedia. Al igual que las taxonomías indicadas en el punto 4, se usan en procesos de clasificación de documentos, pero también en la ingeniería ontológica y en el desarrollo de relaciones semánticas entre entidades.
6) Tratamiento semántico: en este grupo se englobarían diferentes aproximaciones que se caracterizan por estar basadas en los principios y técnicas del web semántico. Se incluirían aquí técnicas como la creación de grafos de categorías, la creación de árboles y esquemas semánticos, la extracción de tripletas, la identificación de relaciones significativas entre términos y su aprovechamiento semántico, etc. Las ontologías, aunque parte integrante del web semántico, se han incluido en un grupo aparte por su importancia.
7) Otros usos: dentro de este grupo pueden englobarse aplicaciones muy específicas o poco representadas numéricamente en los trabajos revisados. Ejemplo de ellos pueden ser definición de perfiles de usuario, patrones de edición colaborativa, o la identificación de eventos y personas.

Con el objetivo de identificar con mayor precisión los temas que han sido objeto de investigación se ha elaborado un listado de términos y expresiones significativas, tomadas de los títulos y resúmenes

revisados, y cuyo contenido se recoge a continuación, en correspondencia con los siete grandes grupos delineados en los párrafos anteriores:

- Recuperación de información: Automatic question generation, Automatic subject induction, Discovering answers, Entity retrieval, Exploratory search, Organizing search results, Query clasiffication, Query expansion for entity, ranking, Query phrase expansion, Semantic question answering, Supervised question classification, Tagging, Wikipedia categories for ad hoc search.
- Procesamiento de entidades: Automatic keyword extraction, Computing word relatedness, Context and keyword extraction, Entity desambiguation, Entity retrieval, Entity semantics, Keyphrase extraction, Matching named entitites, Named entity extraction, Named entity linking, Query expansion for entity ranking, Semantic tags, Wikipedia entity expansion, Word sense desambiguation.
- Indización y clasificación de corpus documentales: Automatic document classification, Automatic document tagging, Categories for document labelling, Conceptual indexing, Corpus building in machine learning, Document clustering, Document context similarity, Document indexing, Document topics, Multimodal document classification, Relevant features for text classification, Tagging, Text categorization.
- Creación y uso de taxonomías: Analysis of cluster structure, Automatic taxonomy extraction, Comparing taxonomies, Derivation of "is a" taxonomy, Method for refining a taxonomy, Taxonomy and clustering, Taxonomy-based information content, Twixonomy, Web taxonomies, Wikipedia category graph.
- Creación y uso de ontologías: Automated construction of domain ontology taxonomies, Automatic ontology generation, CyC ontology, Deriving domain taxonomies, Domain ontological structure, Evolution of ontologies, Extracting ontologies, Mining concepts, Ontological models, Ontology density, Ontology evaluation, Rich ontology extraction, Semi-automatic ontology creation, Topic ontology.
- Tratamiento semántico: Category annotation recommendation, Category graph, Concept graph, Disambiguation of keyword search results on highly heterogeneous structured data, Domain semantic networks, Entity linking, Knowledge trees, Ontology density, Ranking related entities, Semantic knowledge base, Semantic knowledge extraction, Semantic relatedness, Semantic relationships extraction, Semantic resource extraction, Semantic tagging, Semantically related category hierarchies, Triples extraction.
- Otros usos: Automatic blog classification, Automatic mapping of Wikipedia categories, Automatic thesaurus generation, Concept hierarchies, Dbpedia, Dynamic facet hierarchy construction, Generation of dictionaries, Multilingual domain specific resources, Semantic interest profiles, Semantic knowledge base, Semantic recommender, Semantic tags, Terminology, Wikipedia categories clustering, WordNet, YAGO-NAGA.

Si bien la variedad de términos y expresiones utilizados permite afirmar y reforzar los resultados indicados en la revisión cualitativa, sin embargo pone de manifiesto un problema subyacente a las revisiones sistemáticas, como es la disparidad de criterio de los autores en la redacción de títulos, resúmenes y em la selección de palabras clave. Incluso en algunos casos puede detectarse el uso como sinónimos de términos o expresiones que no lo son. En lo referido a la clasificación e identificación del

contenido de los trabajos co una mayor precisión, pone de manifiesto que los investigadores recurren a enfoques mixtos y combinan métodos y técnicas, lo que dificulta una aproximación tradicional, y requiere métodos de procesamiento automático de la información para obtener mejores resultados.

CONCLUSIONES

Debe admitirse que Wikipedia está teniendo una importante influencia en la forma en la que los usuarios se aproximan a la resolución de sus problemas de información, y con notable influencia positiva (Fallis, 2008). La investigación científica no resulta tampoco ajena a la importancia de este recurso de información (Tomaszewski y MacDonald, 2016) A ello no resulta ajena la investigación en organización y recuperación de información, que encuentra en Wikipedia un banco de pruebas de gran valor (Mehdi et alii, 2017).

La primera conclusión que se extrae del estudio realizado es que Wikipedia es un objeto de investigación de interés para diferentes enfoques de la investigación sobre sistemas y técnicas de representación, organización y recuperación de información, tanto en lo que concierne a sus aspectos internos, como al uso externo de sus datos en otras áreas y enfoques de investigación. En segundo lugar, el sistema de categorías de Wikipedia sirve para analizar la evolución socio-temporal de los procesos de organización del conocimiento en entornos colaborativos. En tercer lugar, hay que destacar su uso como herramienta de apoyo y validación en diferentes tipos de aproximaciones al estudio y análisis de corpus documentales.

Al tratarse los artículos de Wikipedia y su sistema de categorías de un corpus documental en continua evolución, cabe señalar que los resultados obtenidos en los diferentes estudios pueden variar, a medio o largo plazo, en virtud del desarrollo de factores externos e internos a la propia enciclopedia. Sirvan como referencia los sistemas de etiquetado social, objeto de gran interés durante la pasada década, que han ido desapareciendo progresivamente de los intereses de la investigación sobre organización del conocimiento.

Finalmente, cabe destacar el potencial que ofrece el sistema de categorías de Wikipedia, en cuanto se trata de un esquema de clasificación universal desarrollado de forma colaborativa, que se asemeja a un tesauro, y que se contrapone a los esquemas de clasificación especializados elaborados en contextos cerrados. Ello ofrece un amplio campo tanto para la validación o la comparación entre esquemas de clasificación existentes, como para creación de otros nuevos desde una perspectiva que permita combinar ambas aproximaciones a la organización del conocimiento. Varios trabajos recientes inciden en esta cuestión (Salah, Gao, Suchecki y Scharnhorst, 2012; Kiyota et alii, 2009), así como en las ventajas y avances del sistema de Wikipedia frente a clasificaciones tradicionales (Jiménez-Pelayo, 2009).

REFERENCIAS BIBLIOGRÁFICAS

Capocci, A., Rao, F., & Caldarelli, G. (2008). Taxonomy and clustering in collaborative systems: The case of the on-line encyclopedia Wikipedia. *EPL (Europhysics Letters)*, *81*(2), 28006. https://doi.org/10.1209/0295-5075/81/28006

Fallis, D. (2008). Toward a Epistemology of Wikipedia. *Journal of the American Society for Information Science and Technology*, *59*(10), 1662-74. DOI 10.1002/asi.20870

Jiménez-Pelayo, J. (2009). Wikipedia como vocabulario controlado: ¿está superado el control de autoridades tradicional?. *El Profesional de la Información, 18*(2), 188-201. DOI 10.3145/epi.2009.mar.09.

Kiyota, Y., Nakagawa, H., Sakai, S., Mori, T., & Masuda, H. (2009). Exploitation of the Wikipedia category system for enhancing the value of LCSH. En *2009 ACM/IEEE Joint Conference on Digital Libraries, JCDL'09* (p. 411). https://doi.org/10.1145/1555400.1555488

Mehdi, M., Okoli, C., Mesgari, M., Nielsen, F. Å., & Lanamäki, A. (2017). Excavating the mother lode of human-generated text: A systematic review of research that uses the wikipedia corpus. *Information Processing & Management*, *53*(2), 505–529. https://doi.org/10.1016/j.ipm.2016.07.003

Mesgari, M., Okoli, C., Mehdi, M., Nielsen, F.A., & Lanamäki, A. (2015). "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*, *66*(2), 219-245.

Nielsen, F.A. (2017). *Wikipedia research and tools: Review and comments*. Working paper. http://www2.compute.dtu.dk/pubdb/views/edoc_download.php/6012/pdf/imm6012.pdf

Okoli, C., & Schabram, K. (2009). Protocol for a Systematic Literature Review of Research on the Wikipedia. En *MEDES '09 Proceedings of the International Conference on Management of Emergent Digital EcoSystems*, Art. 74. https://doi.org/10.1145/1643823.1643912

Salah, A. A., Gao, C., Suchecki, K., & Scharnhorst, A. (2012). Need to Categorize: A Comparative Look at the Categories of Universal Decimal Classification System and Wikipedia. *Leonardo*, 45(1), 84–85. https://doi.org/10.2307/41421810

Smiraglia, R.P., & Cai, X. (2017). Tracking the Evolution of Clustering, Machine Learning, Automatic Indexing and Automatic Classification in Knowledge Organization. *Knowledge Organization*, 44(3), 215-233.

Thornton, K., & McDonald, D.W. (2012). Tagging Wikipedia: collaboratively creating a category system. En *Proceedings of the 17th ACM International Conference on Supporting Group Work*, 219-228. doi: 10.1145/2389176.2389210.

Tomaszewski, R., & MacDonald, K.I. (2016). A Study of Citations to Wikipedia in Scholarly Publications. *Science & Technology Libraries*, 35(3), 246–261. https://doi.org/10.1080/0194262X.2016.1206052

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

Tramullas, J., Garrido-Picazo, P., & Sánchez-Casabón, A.I. (2016). Research on Wikipedia Vandalism: a brief literature review. En *CERI '16 Proceedings of the 4th Spanish Conference on Information Retrieval*, Art. 15. DOI 10.1145/2934732.2934748

Vrandecic, D. (2013). The rise of wikidata. *IEEE Intelligent Systems*, 28(4): 90–95. DOI 10.1109/MIS.2013.119