

Documento de trabajo

Vocabulario de estudios y prácticas editoriales

Wilson Colmenares
Pablo Estrada
Investigadores

Instituto Caro y Cuervo
Grupo de investigación: Literatura
Línea de investigación: El libro en Colombia: prácticas materiales e intelectuales
Observatorio Editorial Colombiano OEC
Sitio web: <https://oec.caroycuervo.gov.co>
Correo electrónico: oec@caroycuervo.gov.co

Fecha de inicio: 25/05/2022
Fecha de culminación: 19/07/2022

Bogotá

Versión: 19/07/2022

Licencia de publicación:



Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional (CC BY-NC-SA 4.0)
<https://creativecommons.org/licenses/by-nc-sa/4.0/deed.es>

Contenido

1. Descripción del problema	3
2. Antecedentes	5
3. Objetivo.....	6
4. Metodología	6
4.1. Fase 1. Diseño e implementación de la metodología	6
4.2. Fase 2. Definición del alcance del vocabulario	6
4.3. Fase 3. Recopilación, selección y sistematización de fuentes de información.....	7
4.4. Fase 4. Publicación del vocabulario en Wikidata.....	9
5. Productos.....	9
6. Grupo de trabajo.....	9
7. Cronograma.....	10

1. Descripción del problema

El campo editorial requiere que se fortalezca la producción de conocimiento que describa las dinámicas, las relaciones y las tensiones de los estudios y de la industria editorial mediante información confiable, periódica y comparable. La dinámica cambiante de la cultura a nivel local, nacional, regional y global exige analizar el fenómeno editorial mediante dimensiones, categorías y variables que a través de metodologías e instrumentos de investigación produzcan fuentes de información primarias y secundarias provenientes de muestras empíricas representativas y particulares que permitan identificar y explicar las recurrencias y las correlaciones constitutivas del campo de estudio.

Para estudiar un campo de conocimiento se requiere de un vocabulario terminológico que permita identificar las estructuras conceptuales, técnicas y sociales del campo, por lo tanto, una de las posibilidades para interpretar, describir y comprender los estudios editoriales es a partir de una exploración conceptual que permita identificar y establecer la terminología propia del campo editorial en la experiencia tanto lingüística como comportamental. Tal acción básica permitirá abordar la materia, interpretar lo acontecido en su contexto y aportará una perspectiva más amplia. Al respecto, Koselleck afirma que “no hay experiencias sin conceptos y no hay conceptos sin experiencias.” (2012, p. 29). Lo que justifica la decisión de iniciar el diseño y la implementación de un vocabulario terminológico controlado de los estudios editoriales como uno de los principales instrumentos colaborativos y tecnológicos que le brinde los fundamentos conceptuales al campo editorial.

“El concepto de vocabulario controlado (lenguaje documental en cualquiera de sus formas) lleva años sometido a intensas tensiones debido a las nuevas formas de organización del conocimiento en sistemas digitales y su adaptación a nuevas funciones como la navegación web, el descubrimiento de recursos o el posicionamiento en buscadores (Lambe, 2007; Pérez-Montoro; Codina, 2017).”¹

“La construcción de un vocabulario controlado debe ser un proceso colaborativo muy dinámico entre sus diferentes redacciones y profesionales, caracterizado además por una gran necesidad de respuesta a nuevos términos y fenómenos sociales. Los conceptos estables han de convivir con otros que surgen al calor de los acontecimientos. Este vocabulario, referido por sus responsables como “árbol de conocimiento”, es un ejemplo de sistema de organización del conocimiento que es en parte un tesoro, una folksonomía interna y que tiende a querer adoptar forma de ontología. Este sistema de etiquetado podría ser entendido como un “organic thesaurus”, conforme a la matriz de “infraestructuras de conocimiento”².

“El proceso de construcción de tesauros está bien estudiado por varios autores clásicos (Broughton, 2006). Es un proceso exigente y prolongado en el tiempo, que tiende a conformar

¹ María José Baños Moreno, Juan Antonio Pastor Sánchez, y Tomás Saorín Pérez, «Uso de Wikidata y Wikipedia para la generación asistida de un vocabulario estructurado multilingüe sobre la pandemia de Covid-19», *El profesional de la información* 29, n.º 5 (2020): 3.

² Baños Moreno, Pastor Sánchez, y Saorín Pérez, 3.

un instrumento de normalización terminológica en el que la estabilidad pesa más que los procesos de revisión y actualización. Más allá del conjunto básico de relaciones jerárquicas y asociativas entre conceptos en la actualidad se tiende hacia una “ontologización” que haga disponibles más tipos de relaciones entre conceptos. Para ello se procede a la formalización en tipos, subclases, propiedades y subpropiedades, lo que, combinado con la correspondencia entre conceptos de diferentes esquemas, permite abordar con mayores expectativas la organización y navegación entre contenidos. Tal como afirma Stuart (2016): para entender mejor el proceso de construcción de vocabularios controlados se deben considerar también técnicas de construcción de ontologías.”³

“En la construcción de un tesoro o de una ontología, en cualquier de las diferentes propuestas metodológicas habitualmente aplicadas, se contemplan aspectos de recolección y selección terminológica y organización de relaciones. Estas tareas requieren una comprensión clara del dominio abordado y una alta cualificación.”⁴

¿Cómo recopilar, sistematizar y organizar la extensa red de conceptos de los estudios editoriales? Y, sobre todo, ¿cómo hacerlo para que el proyecto sea viable, sostenible y escalable en el tiempo?

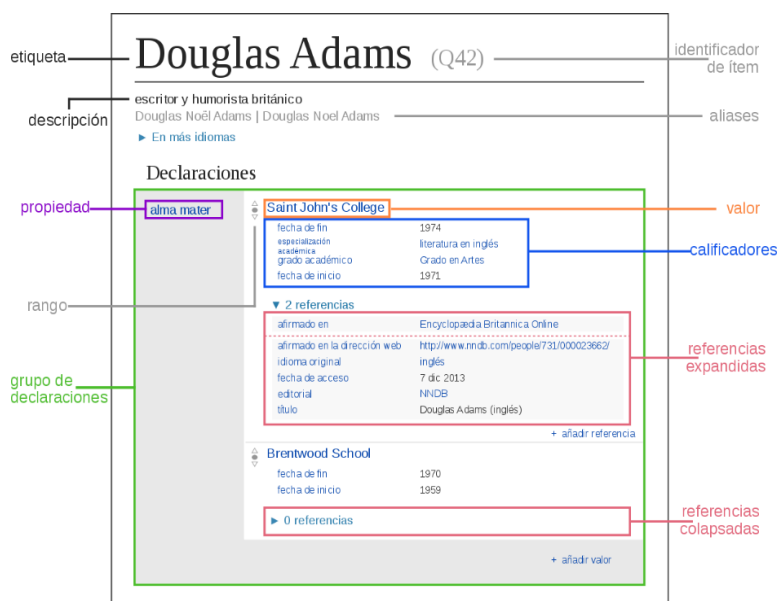
Wikidata y Wikipedia son proyectos tecnológicos y colaborativos que se han convertido en una de las principales fuentes de organización del conocimiento y vocabulario controlado, con contenidos semiestructurados, fácilmente accesibles para su procesamiento con diferentes tecnologías y enfoques. Wikipedia es un recurso utilizado habitualmente en las investigaciones del campo del procesamiento del lenguaje natural (PLN), especialmente para el reconocimiento, contextualización y categorización de entidades (Ehrmann; Rosset, 2016). Esta situación adquiere una dimensión nueva a partir de la maduración del proyecto Wikidata. Wikipedia ofrece artículos textuales, en un proceso continuo de mejora y revisión editorial. En cada una de las más de 200 ediciones en idiomas distintos, los artículos son contenidos independientes, con extensión, calidad, y editores diferenciados. Wikidata amplifica las cualidades de Wikipedia como fuente de “conocimientos conectados” agrupando cada artículo equivalente de cada una de las Wikipedias en una entidad específica con un identificador único, independiente del idioma. Las propiedades de cada uno de los ítems de Wikidata se describen siguiendo un modelo de datos semántico. Los ítems se organizan en tipos y clases, conformando una ontología colaborativa sui generis (Piscopo; Simperl, 2018). Mientras en Wikipedia la edición en cada idioma es un proyecto independiente que elabora y crea sus artículos desde el punto de vista válido para su comunidad de editores, en Wikidata se da forma a un banco de datos central reutilizable por todos los proyectos Wikimedia. Dichas descripciones (claims) adoptan la forma de sujeto-predicado-objeto de manera muy similar a las tripletas RDF. Wikidata podría entenderse como una plataforma de publicación de datos estructurados, con una arquitectura autosuficiente, basada en la creación de declaraciones. El modelo de datos de Wikidata dispone de una representación RDF de manera que los datos son interrogables mediante

³ Baños Moreno, Pastor Sánchez, y Saorín Pérez, 3.

⁴ Baños Moreno, Pastor Sánchez, y Saorín Pérez, 3.

Sparql utilizando Wikidata Query Service (WDQS). Mediante consultas Sparql es posible recuperar las relaciones de un ítem con otros (incluyendo tipos y clases). Los contenidos colaborativos de Wikidata y Wikipedia podrían aprovecharse para crear un sistema de organización del conocimiento o un vocabulario controlado sobre los estudios editoriales mediante métodos y aplicaciones de procesamiento asequibles.⁵

Ontología de información de Wikidata



<https://www.wikidata.org/wiki/Wikidata:Glossary/es>

2. Antecedentes

- Breve glosario multilingüe sobre edición, producción editorial, derechos y contratos: <https://silo.tips/download/breve-glosario-trilingue-sobre-edicion-produccion-editorial-derechos-y-contratos>
- Glosario de edición y diagramación: <https://www.autoreseditores.com/es-co/guia-glosario.html>
- Glosario terminología de edición: <https://silepsiscreaciones.wordpress.com/glosario-terminologia-de-edicion>
- Glosario de términos editoriales: <https://kitzalet.com/editorial-digital/glosario-de-terminos-editoriales-descargable>

⁵ Baños Moreno, Pastor Sánchez, y Saorín Pérez, 6.

- Glosario. Tipografía & producción editorial: <http://www.libros.unam.mx/glosario-tipografia-produccion-editorial-9786073023320-libro.html>
- Latindex: evaluación de revistas electrónicas: <http://vocabularios.caicyt.gov.ar/latindexdig/index.php>
- Wikiglosarios: <https://es.wikipedia.org/wiki/Categor%C3%ADa:Anexos:Glosarios>
- Wikcionario: <https://es.wiktionary.org/wiki/Wikcionario:Portada>

3. Objetivo

Construir un vocabulario estructurado de estudios y prácticas editoriales que permita identificar las estructuras conceptuales, técnicas y sociales a través de un sistema de información colaborativo, sostenible, escalable, accesible e interoperable.

4. Metodología

Se propone un método que permita de forma semiautomática, a partir del enfoque de vocabularios y ontologías terminológicas en el contexto de los datos enlazados construir un vocabulario de los estudios editoriales, a través de herramientas tecnológicas abiertas y colaborativas que permitan recopilar términos y definiciones de fuentes de información y de Wikidata y Wikipedia. El proceso se abordaría desde un punto de vista científico, cultural y socioeconómico. Además, deberá permitir recopilar los conceptos que se producen y registran constantemente en diferentes tipos de publicaciones a través de un procedimiento de validación y generalización del proceso. Por este motivo, este trabajo se apoya en la propia organización y producción de conocimiento presente en el proyecto de contenido colaborativo de Wikidata y de Wikipedia. Teniendo en cuenta que estos proyectos son una fuente de organización del conocimiento y vocabulario controlado, con contenidos semiestructurados, fácilmente accesibles para su procesamiento con diferentes tecnologías y enfoques.

4.1. Fase 1. Diseño e implementación de la metodología

Actividades:

- 1) Recopilar y analizar fuentes de información especializadas.
- 2) Formular e implementar la metodología.

4.2. Fase 2. Definición del alcance del vocabulario

Actividades:

- 3) Hacer una lista con los términos principales.

Vocabulario de estudios y prácticas editoriales

libro
imprensa
impresión
publicación
edición
editorial
revista
periódico

- 4) Hacer una taxonomía con los términos de control.

La taxonomía será elaborada después de haber realizado las actividades 6, 7 y 8 del proyecto.

- 5) Definir el idioma del vocabulario terminológico.

Español, extranjerismos y localismos, según el uso.

4.3. Fase 3. Recopilación, selección y sistematización de fuentes de información

Actividades:

- 6) Buscar, seleccionar, almacenar en digital y hacer una lista con los diccionarios, glosarios, enciclopedias, tesauros y demás fuentes referenciales especializadas en edición, artes gráficas, producción editorial, libro, revistas, etc.

- Diccionarios y glosarios. ([En PDF](#))

- En línea:

Glosario de IberLibro. (2021, agosto 6). Iberlibro (ES).
<https://www.iberlibro.com/libros-antiguos-raroscoleccion/guia-coleccionista-libros/glosario.shtml>

Glosario de términos bibliotecológicos | Observatorio Tecnológico. (s. f.).
<http://recursostic.educacion.es/observatorio/web/eu/cajon-de-sastre/38-cajon-de-sastre/198-glosario-de-terminos-bibliotecologicos>

- 7) Hacer una [lista con los términos](#) que requieren definición del [cuestionario](#) de caracterización de las editoriales independientes colombianas 2019.
Objetivo de la lista de términos:

- a) Fortalecer la formulación, comprensión y respuesta de las preguntas de la encuesta de edición independiente colombiana de 2021.
- b) Tener los términos de control para hacer la taxonomía del vocabulario de estudios y prácticas editoriales.

Criterios de selección de los términos:

- a) Términos principales relacionados de forma directa con los estudios y prácticas editoriales o de edición. Este tipo de términos se consideran como: descriptores.
- b) Términos secundarios (sinónimos y cuasisinónimos) relacionados de forma directa con los estudios y prácticas editoriales o de edición. Este tipo de términos se consideran como: descriptores equivalentes.
- c) Términos generales.

Criterios de registro de los términos:

- a) Términos simples.
- b) Términos compuestos con varias palabras.

Procedimiento:

- a) Normalizar la [lista general y específica de términos](#) antes de iniciar la búsqueda de definiciones teniendo en cuenta los anteriores criterios.
- 8) Buscar, seleccionar y registrar las definiciones de la [lista con los términos](#) que requieren definición del [cuestionario](#) de caracterización de las editoriales independientes colombianas 2019.

Procedimiento:

- a) Agregar nuevas fuentes de información en las carpetas compartidas en el caso de tenerlas.
- b) Buscar los términos en las siguientes fuentes de información:
 - Diccionarios y glosarios. ([En PDF](#))
 - En línea:
 - Glosario de IberLibro. (2021, agosto 6). Iberlibro (ES). <https://www.iberlibro.com/libros-antiguos-raroscoleccion/guia-coleccionista-libros/glosario.shtml>
 - Glosario de términos bibliotecológicos | Observatorio Tecnológico. (s. f.). <http://recursostic.educacion.es/observatorio/web/eu/cajon-de-sastre/38-cajon-de-sastre/198-glosario-de-terminos-bibliotecologicos>

- [Manuales y guías.](#)
 - [Legislación.](#)
 - [Normas técnicas.](#)
- c) Registrar las referencias bibliográficas, los términos y las definiciones bibliográficas con las páginas de las fuentes de información en el [Grupo de Zotero del OEC.](#)
- d) Registrar los términos, las definiciones y las referencias bibliográficas con las páginas de las fuentes de información en el documento del [Glosario.](#)

4.4. Fase 4. Publicación del vocabulario en Wikidata

Actividades:

- 9) Crear el proyecto en Wikidata.
- 10) Diseñar el perfil de aplicación de datos en Wikidata.
- 11) Normalizar los datos según el perfil de aplicación de datos en Wikidata.
- 12) Hacer pruebas de migración de los datos.
- 13) Migrar los datos.
- 14) Diseñar los grafos de visualización de los datos.
- 15) Hacer el informe final del proyecto.

5. Productos

- a) Glosario definiciones de la encuesta de edición independiente de 2021.
- b) Términos y definiciones necesarias para hacer la taxonomía del vocabulario de estudios y prácticas editoriales.

6. Grupo de trabajo

Integrante	Actividad
María Angélica Rodríguez	Recopilación, selección y sistematización
Lorena Cano	Recopilación, selección y sistematización
Pablo Estrada	Verificación de la recopilación, selección y sistematización
Wilson Colmenares	Verificación de la recopilación, selección y sistematización

7. Cronograma

Actividad	Mes				
	Feb	Mar	Abr	May	Jun
1. Recopilar y analizar fuentes de información especializadas.	■				
2. Formular e implementar la metodología.	■				
3. Hacer una lista con los conceptos y términos principales.	■				
4. Hacer una taxonomía con los términos de control.	■				
5. Definir el idioma del vocabulario terminológico.	■				
6. Buscar, seleccionar, almacenar en digital y hacer una lista con los diccionarios, glosarios, enciclopedias, tesauros y demás fuentes referenciales especializadas en edición, artes gráficas, producción editorial, libro, revistas, etc..		■	■	■	■
7. Hacer una lista con los términos que requieren definición del cuestionario de caracterización de las editoriales independientes colombianas 2019.		■	■	■	■
8. Buscar, seleccionar y registrar las definiciones de la lista con los términos que requieren definición del cuestionario de caracterización de las editoriales independientes colombianas 2019.		■	■	■	■
9. Crear el proyecto en Wikidata.					■
10. Diseñar el perfil de aplicación de datos en Wikidata.					■
11. Normalizar los datos según el perfil de aplicación de datos en Wikidata.					■
12. Hacer pruebas de migración de los datos.					■
13. Migrar los datos.					■
14. Diseñar los grafos de visualización de los datos.					■
15. Hacer el informe final del proyecto.					■