

APLICACIONES DE LA INTELIGENCIA ARTIFICIAL EN EL MODELADO DE LA ESTRUCTURA SEMÁNTICA-RELACIONAL DE UN TESAURO DE BIBLIOTECOLOGÍA Y ESTUDIOS DE LA INFORMACIÓN

ADRIANA SUÁREZ SÁNCHEZ*

CATALINA NAUMIS PEÑA**

Resumen: *Este estudio tiene como objetivo evaluar la utilidad/no utilidad de la inteligencia artificial (IA) en la determinación conceptual y relacional de los descriptores que conforman un tesauro. La norma ISO 25964-1: Información y documentación. Parte 1: Tesoros para la recuperación de información destaca la necesidad de vincular los términos de un tesauro con los conceptos que les subyacen. La determinación conceptual también cumple un papel clave en la redacción de notas de alcance (NA) y en la definición de estructuras asociativas entre los descriptores, incluyendo relaciones de equivalencia (USE), jerárquicas (término general [TG], término específico [TE]) y asociativas (término relacionado [TR]). A partir de los procedimientos que realiza un profesional humano, se diseñaron tres instrucciones para guiar a la IA en la construcción de notas de alcance, relaciones de equivalencia y nodos semánticos. Los resultados de estas simulaciones son posteriormente analizados y discutidos.*

Palabras clave: *Tesoros; Inteligencia Artificial; Organización del Conocimiento.*

Abstract: *This study aims to evaluate the usefulness/non-usefulness of artificial intelligence (AI) in the conceptual and relational determination of the descriptors that make up a thesaurus. ISO 25964-1: Information and documentation. Part 1: Thesauri for information retrieval emphasizes the need to link the terms in a thesaurus with the concepts underlying them. Conceptual determination also plays a key role in the writing of scope notes (SN) and in the definition of associative structures between descriptors, including equivalence (USE), hierarchical (broader term [BT], narrower term [NT]) and associative (related term [RT]) relationships. Based on the procedures performed by a human professional, three instructions were designed to guide the AI in the construction of scope notes, equivalence relations and semantic nodes. The results of these simulations are subsequently analyzed and discussed.*

Keywords: *Thesauri; Artificial Intelligence; Knowledge Organization.*

* Instituto de Investigaciones Bibliotecológicas y de la Información, Universidad Nacional Autónoma de México (UNAM). Email: asuarez@iibi.unam.mx. ORCID: <https://orcid.org/0000-0003-0215-337X>.

** Instituto de Investigaciones Bibliotecológicas y de la Información, Universidad Nacional Autónoma de México (UNAM). Email: naumis@unam.mx. ORCID: <https://orcid.org/0000-0003-3152-3958>.

INTRODUCCIÓN

El diseño de tesauros ha sido una tarea continua desde la segunda mitad del siglo XX, impulsada por la expansión y el surgimiento de diversos campos del conocimiento con una terminología especializada. Los tesauros son aplicados en la indización temática de recursos de información en bibliotecas físicas y digitales, así como en archivos, museos, galerías y otras unidades de memoria. De igual modo, en el entorno digital, son lenguajes de indización que controlan el vocabulario y facilitan la recuperación de información en sitios gubernamentales, ámbitos educativos, plataformas de entretenimiento y portales comerciales.

El World Wide Web Consortium (W3C) ha destacado la importancia de los tesauros en la web semántica en el documento oficial del W3C sobre SKOS (Simple Knowledge Organization System), un modelo de datos que facilita la interoperabilidad y una recomendación del W3C para diseñar tesauros y otros sistemas de organización del conocimiento en la Web Semántica.

Tesauros, esquemas de clasificación, listas de encabezamientos de materias, taxonomías y otros tipos de vocabularios controlados son ejemplos de sistemas de organización del conocimiento que pueden beneficiarse de una representación en un formato estándar y procesable por máquina. El Sistema Simple de Organización del Conocimiento (SKOS) ofrece un modelo para expresar estos sistemas en el marco de la Web Semántica (Miles y Bechhofer 2009, p. 1).

La construcción de tesauros atañe a los bibliotecólogos, lingüistas, profesionales de la información y numerosos expertos en organización del conocimiento y recuperación de recursos de información. Es una tarea evolutiva que día a día se nutre de numerosas teorías y se apoya en instrumentos tecnológicos. Hacia finales del siglo pasado las tecnologías asociadas al Procesamiento del Lenguaje Natural (PNL) aportaron enormes ventajas en la construcción de tesauros y, en tiempos más recientes, los avances en las tecnologías de la información y el conocimiento, especialmente en el campo de la Inteligencia Artificial, han impulsado un replanteamiento sobre su utilidad en diversos procesos y actividades. El servicio del tesoro es ofrecer descriptores o términos normalizados en un área del conocimiento, que son aptos para resumir contenidos documentales y están relacionados entre sí para conformar una red de significados que enmarquen y otorguen sentido a términos aislados del discurso.

Dado que la IA puede automatizar procesos, analizar datos, definir significados y resolver problemas, el presente artículo tiene por objetivo evaluar la utilidad de ChatGPT (OpenAI 2025) en la determinación semántica y relacional de los descriptores que conforman un tesoro sobre Bibliotecología y Estudios de la información que

se construye en el Instituto de Investigaciones Bibliotecológicas y de la Información de la Universidad Nacional Autónoma de México.

El supuesto de la investigación estableció que la IA puede apoyar en muchas de las actividades necesarias para desarrollar un tesauro. En cuanto a la determinación de la estructura conceptual, se asumió que la IA posee la capacidad de consultar corpus y bases de datos actualizadas y diversas que le permitan identificar conceptos sobre un descriptor. En lo tocante a la estructura relacional característica de los tesauros, se asumió que la IA generativa registra el potencial necesario para realizar inferencias sobre los vínculos semánticos posibles establecer entre los descriptores de un campo de conocimiento. Por último, es importante mencionar que el alcance experimental del artículo consideró un corpus reducido que evidencia, de manera general, las ventajas/desventajas de la IA en los objetivos establecidos.

1. MARCO TEÓRICO

Los tesauros son un sistema de organización del conocimiento altamente especializado que emergió pasada la primera mitad del siglo XX, con el propósito de representar temáticamente campos disciplinares y organizar recursos de información sobre tales temas. Como sistemas para la organización del conocimiento o vocabularios controlados, aportan esos elementos que contextualizan los términos al presentar una estructura del conocimiento en jerarquías, validación de sinónimos, relaciones de proximidad con otros términos, explicación del uso del término o significado de este.

Todo proceso enfocado a la planeación y desarrollo de tesauros supone el estudio y definición de la situación comunicativa para presentar estructuras que le otorguen validez a cada uno de los elementos léxicos incluidos en el vocabulario controlado. En este sentido, su construcción implica analizar cada término, sus sinónimos, las clases y las relaciones que deben ser revisadas y determinadas, con miras a su integración en una estructura sistémica, cuyo objetivo es representar, con el mayor grado de fidelidad posible, un dominio de conocimiento.

En palabras de Currás (2010, p 119), la sistematicidad contempla:

- Un sistema principal: el tesauro en sí.
- Un subsistema primario: los términos en sí mismos.
- Un subsistema secundario: las clases de términos.
- Un subsistema terciario: las relaciones entre los términos en un orden jerárquico.

Las etapas de determinación conceptual y establecimiento de relaciones son esenciales en la construcción de tesauros. Las normas *ISO: Information and documentation. Part 1: Thesauri for information retrieval* (International Organization for Standardization 2011) e *ISO 25964-2:2013. Part 2. Interoperability with other*

vocabularies (International Organization for Standardization 2013) consideran necesario vincular los términos de un tesoro con los conceptos que le subyacen. Aunado a ello, la determinación conceptual es de suma utilidad en el establecimiento de notas de alcance (NA) y en el asentamiento de asociaciones entre los descriptores de un tesoro (relaciones de equivalencia (USE), relaciones jerárquicas (TG Término general y TE Término específico) y relaciones asociativas (TR Término relacionado).

Durante casi todo el siglo XX, el modelado semántico y relacional de los tesauros se realizó de forma manual. La tarea consistía en asociar el término con un concepto, la identificación de términos sinónimos y el establecimiento de relaciones entre el descriptor y su término más amplio (TG), su término específico (TE) y sus términos relacionados (TR). Tal proceso implicó gran esfuerzo humano y largas jornadas de trabajo en las que se preparaban tarjetas con los datos de cada término hechas a mano. Los desarrolladores de tesauros frecuentemente señalaban que la actividad era ardua y complicada debido a la desactualización de diccionarios, enciclopedias, libros u otras fuentes de información que apoyaran la actividad,

Hacia finales del siglo XX las tecnologías de la información trasformaron el panorama de los vocabularios controlados. Emergieron numerosas aplicaciones computacionales para el Procesamiento del Lenguaje Natural que hicieron posible la extracción terminológica automatizada, la identificación de contextos definitorios en corpus textuales, el diseño de redes de copalabras y la construcción de nodos semánticos.

Aunado a las tecnologías de PLN antes mencionadas, en los últimos años la Inteligencia Artificial (IA) se vislumbra como una herramienta auxiliar para el desarrollo de tesauros. Al hablar de IA estamos ante una aplicación computacional que promete la automatización de muchas tareas, pero sobre la que persisten incontables dudas. Por Inteligencia Artificial debemos entender «un objeto o entidad artificial que posee la capacidad de cumplir o superar los requisitos de la tarea asignada, considerando las circunstancias culturales y demográficas» (Kelly, Kaye y Oviedo-Trespacios 2023, p. 2). La IA «considera desde un conjunto de algoritmos hasta sistemas que muestran un comportamiento inteligente al analizar su entorno y tomar medidas, con cierto grado de autonomía, para lograr objetivos específicos» (Sheikh, Prins y Schrijvers 2023, p. 15).

Otros autores han su capacidad de ofrecer soluciones y la flexibilidad para adaptarse al entorno.

la IA es una tecnología que, apoyándose en aportes de gran cantidad de datos o en un corpus documental específico, es capaz de desarrollar unos programas y métodos que permiten imitar el razonamiento inteligente [...] Nos referimos a sistemas que, a partir de la información que tienen y su análisis, son capaces de encontrar soluciones a problemas complejos o desarrollar nuevas funciones que le son requeridas (Martínez de Madariaga 2023, p. 93).

Las aplicaciones de IA en el desarrollo de vocabularios controlados cuentan con vasta producción académica en el rubro de los diccionarios, donde las impresiones son diversas. Mientras De Schryver (2023, p. 355) señala que «ha comenzado una nueva era, la de la aplicación exitosa de la IA en la lexicografía». Vossen (2022, párr. 1) asume que el empleo de IA en el desarrollo de vocabularios es «una pérdida de tiempo». Pese a los puntos de vista encontrados, en general se destaca la utilidad de la IA en procesos de obtención terminológica, conceptualización, detección de sinónimos, identificación de palabras en contexto, extracción de nodos semánticos, entre más.

En cuanto a la utilidad de la IA en el desarrollo de tesauros, las investigaciones son casi nulas. Al realizar una búsqueda exploratoria, en Scopus se recuperan cuatro estudios y tres en Web of Science en el rango temporal 2024-2025. Aunque los artículos declaran la relación Inteligencia Artificial-Tesauros en sus palabras clave, no tratan sobre aplicaciones específicas de IA generativa en la construcción de tesauros. A lo sumo, de manera teórica, retoman sus ventajas en la obtención de términos de un campo de conocimiento o la posibilidad de establecer definiciones entorno a un descriptor. Pese a la poca investigación sobre el tema, los artículos coinciden en dos aspectos: primero, sólo la experimentación podrá arrojar luces sobre el asunto y, segundo, que entre las diversas actividades que los humanos llevan a cabo para desarrollar un tesauro, en algunas la IA puede obtener mejores resultados que en otras.

2. MÉTODOS

El estudio se caracterizó por una metodología exploratoria-descriptiva, con rasgos mixtos.

Como técnicas de investigación se emplearon:

1. Hermenéutica del discurso: es el «arte de interpretar textos» (Beuchot 2008, p. 32). A partir de un conjunto establecido de recursos de información se llevaron a cabo procesos de análisis e interpretación que establecieron, de manera crítica, significados o perspectivas sobre el tema.
2. Experimentación aplicada: emplea métodos prácticos para descubrir aspectos de un tema. Es ampliamente utilizada en la industria, el sector médico y la evaluación de tecnologías. Permite al acercamiento exploratorio a un problema. Aplica preceptos de ensayo-error-reformulación.

Mediante la hermenéutica del discurso se revisaron los aspectos teóricos sobre IA y desarrollo de tesauros. Luego, a través de la experimentación aplicada se analizaron las posibilidades de la IA para llevar a cabo tareas de definición léxica y diseño de la estructura relacional de los descriptores de un tesauro. Se trabajó con la IA ChatGPT disponible de forma gratuita para el público en general y con una versión de paga que ofrece mayor tiempo de uso.

ChatGPT (Chat Generative Pre-Trained) fue lanzada al público el 30 de noviembre de 2022. Es una aplicación de IA que simula conversaciones con el usuario. Constituye un modelo de inteligencia artificial capaz de responder preguntas, detectar datos, formular previsiones, redactar textos, entre otras funciones. Se fundamenta en un sistema de procesamiento de lenguaje natural. Emplea una arquitectura de aprendizaje profundo y se nutre de grandes volúmenes de datos en la web que le permiten generar respuestas adecuadas en una amplia variedad de consultas y solicitudes.

Teniendo como base el trabajo que realiza un profesional humano en la determinación semántica y relacional de los descriptores de un tesoro, se consultó a la IA para que construyera la estructura semántica y relacional de algunos descriptores del tesoro. Cada uno de los elementos sobre los que se interrogó a la IA quedó definido del modo siguiente:

- Notas de alcance (NA): son empleadas para conceptualizar un descriptor en un contexto específico. Algunas veces trazan límites sobre una restricción o alcance significativo, mientras que otras veces establecen precisiones de uso, en función de una región geográfica, una comunidad de usuarios, un periodo temporal, entre otros aspectos.
- Relaciones de equivalencia (USE): son empleadas para redireccionar descriptores con significado equivalente o similar hacia otro descriptor, al cual aluden. Cuando se establece una relación de equivalencia puede ser bajo los rubros sinónimo, cuasi sinónimo o el envío de un descriptor que se ha identificado como no aceptado hacia otro descriptor que ha recibido el rango de aceptado, según un criterio establecido previamente.
- Término genérico (TG): es un descriptor que representa una categoría más amplia. Suele ser superior a nivel jerárquico o semántico. Considera el descriptor, al cual alude, como un tipo de relación inferior.
- Término específico (TE): es un descriptor que representa una categoría más específica. Suele ser inferior a nivel jerárquico o semántico. Considera el descriptor, al cual alude, como un tipo de relación superior.
- Término relacionado (TR): es un descriptor que se vincula a otro, al cual alude, bajo algún tipo de relación semántica. Considera diversos tipos de relaciones: misma categoría, disciplina o campo de estudio, proceso/agente, ocupación/persona que la ejerce, entre más. El estándar más actual de tesauros enuncia once diferentes tipos de TR. También menciona que los desarrolladores de tesauros pueden establecer múltiples tipos de asociaciones, siempre que sean necesarias, importantes o se hagan presentes en el dominio de conocimiento que contempla el tesoro.

Considerando algunos descriptores caso o ejemplo, se interrogó a la IA sobre aspectos significativos y relacionales. Para realizar tal escrutinio, se diseñaron instrucciones (*prompts*) (Tabla 1), siguiendo las recomendaciones de Despujol Zabala (2020) y Gómez Cardosa y Garcia-Brustenga (2023), quienes señalan la importancia de elaborar preguntas claras, concisas y contextualizadas, con verbos precisos que dirijan las acciones de las aplicaciones conversacionales.

Tabla 1. Tres instrucciones construidas para consultar a la IA

ACCIONES	INSTRUCCIÓN
Estructura semántica	
Notas de alcance (NA)	A partir de definiciones que localices del término «X» en artículos académicos, elabora una nota de alcance del término, bajo la etiqueta NA, que evidencie el significado del término en el contexto bibliotecario.
Estructura relacional	
Relaciones de equivalencia (USE)	Identifica sinónimos del término «X» en artículos académicos y registra cada uno bajo la etiqueta USE.
Sinónimos	
Siglas	
Nombre científico	
Término con diferente ortografía	Genera una red semántica para el término «X» que incluya: el término «X» como una entrada principal, un término más amplio bajo la etiqueta TG, un término más específico bajo la etiqueta TE y varios términos asociados, cada uno bajo la etiqueta TR.
Relaciones jerárquicas (TG), (TE)	
Genéricas	
Todo/Parte	
Enumerativa	
Relaciones asociativas (TR)	
Misma categoría	
Disciplina o campo de estudio	
Proceso/Agente	
Ocupación/Persona que la ejerce	
Acción/Producto derivado	
Acción/Sujeto Paciente	
Término/Propiedades	
Término/Origen	
Término/Causa	
Materia prima/Producto	
Acción/Propiedad	

Fuente: Elaboración propia

En la ejecución de las instrucciones se eligieron, de manera aleatoria, cinco descriptores del tesauro: Libros digitales, Organización del conocimiento, ISBD,

Modelos conceptuales y Pódcast. El proceso experimental se realizó con los cinco descriptores antes mencionados; sin embargo, por cuestiones de extensión del presente capítulo sólo se ejemplifica con el descriptor Libros digitales. En la evaluación de los resultados, derivados de las tres interrogaciones, se empleó una escala Likert con cuatro rubros (muy útil, útil, poco útil y nada útil). La evaluación de los resultados fue realizada por tres profesionales, con formación de Doctorado en Bibliotecología y Estudios de la Información, quienes también cuentan con amplia experiencia en el desarrollo de tesauros.

3. ANÁLISIS Y DISCUSIÓN DE LOS RESULTADOS

ChatGPT-4, última versión de la reconocida IA ha sido propuesta como una versión mejorada y generativa que ofrece enormes ventajas en las interrogaciones y tareas asignadas. Pese a todas las promesas de la IA, tras la serie de experimentos emprendidos en la presente investigación se evidenció una actitud optimista a la vez que dubitativa sobre las tareas que realiza en la determinación semántica y relacional de los descriptores de un tesauro.

Las opiniones acerca del comportamiento de la IA son observadas y comentadas en la literatura:

GPT-4 contiene grandes mejoras [...] los investigadores afirman que estas capacidades tienen el potencial de transformar la ciencia, pero algunos se muestran frustrados por no poder acceder aún a la tecnología, a su código subyacente ni a la información sobre cómo se entrenó. Esto genera preocupación sobre la seguridad de la tecnología y la hace menos útil para la investigación, afirman los científicos (Sanderson 2023, p. 1).

En una primera etapa del estudio, se generaron notas de alcance (NA) por cada uno de los descriptores. Las NA delimitan el significado de un término en un contexto específico y orientan a los indizadores sobre el uso correcto del vocablo, según una noción más amplia, restringida o criterios específicos de uso.

Tabla 2. Evaluación de las notas de alcance (NA) generadas con IA

Notas de alcance NA	Libros digitales	Organización del conocimiento	Descripción Bibliográfica Internacional Normalizada	Modelos conceptuales	Pódcast
Resultado	Útil	Útil	Poco Útil	Útil	Útil

Fuente: Elaboración propia

Como se observa (Tabla 2), cuatro consultas (80%) obtuvieron resultados útil y pueden ser parcialmente integradas en el tesauro. Las notas generadas por ChatGPT contienen información correcta; sin embargo, a partir de la instrucción dada, su extensión es demasiado amplia (hasta 500 palabras). Tal hecho contrasta con la teoría de los tesauros, la cual establece que las NA suelen ser explicaciones cortas sobre aspectos preferentes de un término. Un segundo inconveniente observado en las notas de alcance radicó en el desconocimiento de la IA sobre el contexto local o regional de tesauro. Mientras la tarea humana contempla la garantía de uso de cada descriptor en el dominio o la biblioteca en donde se usará el tesauro, la IA construye NA a partir de significados o información disponible en Internet, sin establecer criterios temporales, geográficos o preferencias de uso del vocablo en un contexto particular.

En cuanto a la validez académica de las notas de alcance, al establecer en la instrucción la premisa:

A partir de definiciones que localices del término «X» en artículos académicos, elabora una nota de alcance...

En los resultados ChatGPT evidencia las referencias documentales empleadas directa o indirectamente para construir las notas de alcance. Por demás, también agrega recursos académicos que pueden aportar más información sobre el descriptor.

En el caso Libros digitales se obtuvo:

NA: El término libros digitales se refiere a obras monográficas disponibles en formato electrónico, que pueden ser consultadas mediante dispositivos como computadoras, lectores electrónicos, tabletas o teléfonos móviles. Estos libros mantienen las características estructurales de los impresos, pero ofrecen funcionalidades adicionales como la búsqueda de texto completo, anotaciones y enlaces interactivos. «En el contexto bibliotecario, los libros digitales amplían el acceso a la información, superando barreras temporales y espaciales, y requieren que las bibliotecas adapten sus servicios y tecnologías para su gestión y difusión».

La IA establece que la primera parte de la NA, hasta antes de las comillas, fue construida a partir de sus patrones de razonamiento, derivado de consultar diversas fuentes de información. Las fuentes que menciona como consultadas son 10 referencias que la IA enlaza mediante su Localizador de Recursos Uniforme (URL). Luego, en la parte entrecomillada de la nota de alcance, ChatGPT señala que tal sección ha sido

extraída textualmente de una fuente de información, que cita y vincula, mediante su URL. En el modelado de la respuesta, la IA aplicó razonamiento de máquina y adhesión completa a una fuente de información. Consecuentemente, es posible sintetizar que la nota de alcance generada por la IA cumple con criterios de garantía literaria.

Con miras a mejorar los resultados en el desarrollo de las NA, se sugiere especificar su extensión con una precisión mínima o máxima del número de palabras que deben estar contenidas en la respuesta. En proyectos más formales también se sugiere limitar el conjunto de fuentes de información que se establece como marco de referencia, por ejemplo, cambiar la especificación:

- A partir de definiciones que localices del término «X» en artículos académicos... Sustituyéndola por
- A partir de definiciones que localices del término «X» en artículos académicos de la base de datos «X» o el sitio web «X»...

Con el claro objetivo de mejorar los resultados en este tipo de proyectos aplicados, también es posible construir un sitio web con documentos (artículos, libros, ponencias, etc.) de la especialidad que sirvan como corpus de referencia para la IA. En la presente investigación ChatGPT realizó las tareas a partir de lo que puede consultar en la web abierta. Al crear un sitio web que albergue un corpus de referencia, se integra la posibilidad de contar con fuentes de información altamente especializadas que produzcan resultados más confiables.

En cuanto a las relaciones USE, son marcas de reenvío de un término no aceptado hacia uno aceptado. Los desarrolladores de tesauros detectan términos con significado equivalente y discriminan respecto a cuál será el descriptor aceptado frente a los otros que serán redirigidos. En el Tesauro de Bibliotecología, Archivología e Información ya se contaba con el descriptor aceptado y sólo se interrogó a la IA sobre términos equivalentes (Tabla 3) que podían ser establecidos como sinónimos.

Tabla 3. Evaluación de las relaciones de equivalencia (USE) generadas con IA

USE	Libros digitales	Organización del conocimiento	Descripción Bibliográfica Internacional Normalizada	Modelos conceptuales	Pódcast
Resultado	Poco Útil (obtuvo 6 términos, 3 útiles)	Poco Útil (obtuvo dos términos, 1 útil)	Poco Útil (obtuvo 6 términos, 1 útil)	Útil (obtuvo 5 términos, 3 útiles)	Nada Útil (obtuvo 8 términos, ninguno útil)

Fuente: Elaboración propia

Continuando con el caso ejemplo, en el descriptor Libros digitales se identificaron los siguientes términos equivalentes:

Libros electrónicos	
E-books	
Ciberlibros	USE Libros digitales
Libros en formato digital	
Publicaciones electrónicas	
Textos digitales	

ChatGPT propone 6 términos como sinónimos de Libros digitales. La utilidad de éstos es relativa, mientras *Libros en formato digital* es una evidente opción para un reenvío USE, los términos *E-books* y *Ciberlibros* son empleados de manera indiscriminada como sinónimos de *Libros electrónicos* (Ávila Álvarez 2009) y como sinónimos de *Libros digitales* (García-Delgado Giménez y Arias Rubio 2014). En cuanto a la igualación sinonímica entre *Libros electrónicos* y *Libros digitales* obras como *Online Dictionary for Library and Information Science* (Reitz 2013) validan tal razonamiento. Por último, *Publicaciones electrónicas* y *Textos digitales* no pueden ser considerados sinónimos estrictos o relaciones USE de Libros digitales.

En cuanto a la garantía literaria para proponer tales relaciones de equivalencia, la IA no cita fuentes documentales en las cuales sea posible verificar la sinonimia. Mientras que en las NA los resultados ofrecidos por ChatGPT eran más transparentes, en la generación de relaciones USE se asume que el razonamiento de la IA conllevó a tales respuestas. Ante resultados como estos, los expertos han empezado a hablar de respuestas de caja negra, esto es, una respuesta de la IA que no sabemos cómo fue generada y cuyo producto informativo pueden ser completamente, parcialmente o nada correcto.

En la última parte del experimento ChatGPT desarrolló relaciones jerárquicas entorno a un descriptor (Tabla 4). En la determinación de relaciones la unidad léxica es colocada en un marco contextual que evidencia su posición semántica entre otros descriptores que pertenecen a un mismo dominio. Las jerarquías de los tesauros consideran: el término genérico o más amplio (TG), el término más específico (TE) y los términos relacionados (TR).

La construcción manual de tesauros implica un arduo y complejo trabajo para identificar y establecer cada uno de los elementos antes mencionados. En las relaciones TG y TE cada término es analizado desde una perspectiva jerárquica que puede ser general, todo-parte o enumerativa. En cuanto a las relaciones TR un campo de conocimiento engloba denominaciones que mantienen diversos vínculos: misma categoría, disciplina que los estudia, causa-efecto, materia prima-producto

u otras relaciones que puedan ser consideradas esenciales para establecer el marco semántico de un descriptor.

Tabla 4. Evaluación de las relaciones jerárquicas (TG y TE) y asociativas (RT) generadas con IA

TG, TE, TR	Libros digitales	Organización del conocimiento	Descripción Bibliográfica Internacional Normalizada	Modelos conceptuales	Pódcast
Escala	Muy Útil (obtuvo 7 términos, 7 adecuados)	Muy Útil (obtuvo 8 términos, 7 adecuados)	Útil (obtuvo 9 términos, 6 adecuados)	Muy Útil (obtuvo 9 términos, 8 adecuados)	Poco Útil (obtuvo 10 términos, 6 adecuados)

Fuente: Elaboración propia

Siguiendo el caso ejemplo Libros digitales se obtuvo la siguiente estructura:

Libros digitales
TG: Recursos digitales
TE: Libros digitales interactivos
TR: Lectores de libros digitales
Bibliotecas digitales
Préstamo digital
Digitalización de libros
Derechos digitales

En comparación con los rubros NA y USE, la IA no pudo construir respuestas bajo la premisa:

A partir de la consulta del término «X» en artículos académicos, genera una red semántica para el término «X» que incluya: el término «X» como una entrada principal, un término más amplio bajo la etiqueta TG, un término más específico bajo la etiqueta TE y varios términos asociados, cada uno bajo la etiqueta TR.

Tal hecho bien podría implicar que las fuentes académicas disponibles no contenían elementos para llevar a cabo la tarea; sin embargo, tal aseveración no es comprobable. En comparación, cuando se le solicitó a la IA generar la red semántica sin considerar fuentes estrictamente académicas, sí le resulto posible ejecutar el proceso. Con base en la lógica estructural típica de los tesauros, es posible validar las relaciones existentes entre los términos, pese a ello, no es posible saber qué algoritmos empleó

la IA para llegar a tal respuesta. La IA tampoco evidencia las fuentes de información consultadas en la generación de respuestas. Así pues, no se cuenta con alguna garantía literaria que sustente la estructura categorial derivada.

CONCLUSIONES

La construcción de tesauros es una actividad costosa en términos de tiempo y esfuerzo humano. Las tecnologías de la información, desarrolladas en las tres últimas décadas han facilitado la actividad, cuando son aplicadas con conocimiento y enfoque de acción. Los fundamentos del Procesamiento de Lenguaje Natural aunados a avanzados programas han permitido automatizar el desarrollo de tesauros.

A lo largo del tiempo, las tecnologías han tenido un rol trascendental en la construcción de tesauros. Como señalan Mohsen et al. (2018, p. 244), el acercamiento manual a la construcción de tesauros se observa hoy como una tarea ardua y, actualmente, las tecnologías pueden auxiliar en el proceso. La IA es una innovadora tecnología en la construcción de sistemas para organizar el conocimiento. No obstante, a semejanza de los programas de Procesamiento de Lenguaje natural, la IA es solamente una herramienta auxiliar que no sustituirá la tarea humana, especialmente en los rubros de análisis, discernimiento y precisiones significativas.

En la presente investigación no fue posible diseñar una instrucción con la cual se pudiera generar todo el nodo de un descriptor, esto es, con las etiquetas NA, USE, TG, TE y TR juntas. Aunque se realizaron esfuerzos, las instrucciones resultaron demasiado complejas para la IA y, frecuentemente, la aplicación atendía bien a una parte de la instrucción o bien otra. Sólo al tratar cada sección de manera separada las interrogaciones empezaron a producir resultados.

Se detectó que la IA aún carece de datos suficientes que le proporcionen capacidad de razonamiento. En este sentido, los desarrolladores apuntan sobre la característica generativa de la IA, que poco a poco irá mejorando sus respuestas, a partir de las correcciones de los usuarios y la continua adhesión de corpus que la nutran. Las expectativas al respecto son altas y sólo los tiempos venideros evaluarán mejoras. Por ahora, la IA puede ser de suma utilidad en el desarrollo de tesauros, pero aún requiere evaluación y validación humana, especialmente ante graves errores en las respuestas que pueden estar asociados con información errónea, noticias falsas y problemas de codificación de los datos almacenados en la web. Al trabajar sobre la web general, ChatGPT están expuesta al error y la desinformación.

En la investigación se observó que una enorme limitante para la validez de las respuestas de ChatGPT es la incapacidad de la IA para acceder a recursos de información de acceso cerrado, por ejemplo, los artículos académicos o libros que sólo son objeto de consulta cuando se compra cierta base de datos, las bibliotecas digitales de acceso restringido o los repositorios académicos disponibles con contraseñas de acceso.

Se concluye que el uso de Inteligencia Artificial puede ser una herramienta auxiliar en las dos etapas de desarrollo del tesoro antes señaladas:

- Determinación semántica:
 - Asentamiento: contribuye en la formación de conceptos asociados a un término, a partir de fuentes de información.
 - Detección de sinónimos y variantes léxicas: mediante la consulta de fuentes en Internet puede sugerir equivalencias o términos preferentes.
 - Análisis de coocurrencias: identifica términos que aparecen frecuentemente juntos, ayudando a establecer relaciones semánticas.
- Determinación de relaciones:
 - Agrupación automática de términos: mediante modelos de aprendizaje supervisado y no supervisado puede categorizar términos en clases temáticas o jerárquicas.
 - Generación de relaciones semánticas: a través de algoritmos de redes neuronales pueden aprender patrones de asociación entre conceptos y generar automáticamente relaciones jerárquicas y asociativas.

Aunque ChatGPT puede apoyar en el desarrollo de tesauros, la relevancia de la actividad requiere garantías de confianza, fiabilidad y curación de datos que sirvan como fundamento de las NT, USE, TG, TE o TR. En estudios futuros se recomienda llevar a cabo experimentos con GPTs más específicos, por ejemplo: Scholar GPT, Consensus, SciSpace o utilizar Perplexity con documentos académicos.

Muchas investigaciones sobre la utilidad de la IA en el desarrollo de diccionarios u otros vocabularios controlados han concluido que sería extremadamente peligroso confiar en un gran modelo lingüístico que no permitiera el acceso a los datos subyacentes en los que se basa su resultado. Tal precepto es aplicable al desarrollo de tesauros en la bibliotecología; mientras el recurso humano está en posibilidad de argumentar sobre el proceso realizado, la examinación de los datos y evidenciar el corpus que afianza el tesoro, en muchas de las respuestas la IA no señala fuentes y el proceso llevado a cabo se mantiene oculto al usuario, en esa sección que ha sido denominada caja negra de la IA.

REFERENCIAS

- ÁVILA ÁLVAREZ, Antonio María, 2009. *El libro electrónico* [En línea]. Madrid: Boletín Económico de ICE [consult. 2025-06-18]. Disponible en: <https://www.revistasice.com/index.php/BICE/article/view/4835>.
- BEUCHOT, Mauricio, 2008. *Perfiles esenciales de la hermenéutica*. México: Universidad Nacional Autónoma de México: Fondo de Cultura Económica.

- CURRÁS, Emilia, 2010. *Ontologies, taxonomies and thesauri in systems science and systematics*. Reino Unido: Chandos.
- DE SCHRYVER, Gilles-Maurice, 2023. Generative AI and lexicography: the current state of the art using ChatGPT. *International Journal of Lexicography* [En línea]. **36**(4), 355-387 [consult. 2025-06-18]. Disponible en: <https://doi.org/10.1093/ijl/ecad021>.
- DESPUJOL ZABALA, Ignacio, 2020. *Uso de ChatGPT. Consejos para hacer prompts* [En línea] [consult. 2025-06-18]. Disponible en: <https://www.upv.es/visor/media/3ab8f200-ffc6-11ed-b888-d3fdbf7baf29/v>.
- GARCÍA-DELGADO GIMÉNEZ, Belén, y Gala ARIAS RUBIO, 2014. Papel vs. digital: hábitos de lectura de los estudiantes de la UEM. *Caracteres: estudios culturales y críticos de la esfera digital* [En línea]. **3**(1), 1-22 [consult. 2025-06-18]. Disponible en: <http://revistacaracteres.net/revista/vol3n1mayo2014/papel-digital-lectura-estudiantes>.
- GÓMEZ CARDOSA, Desirée, y Guillem GARCIA-BRUSTENGA, 2023. *¿Cómo preguntar a la IA? Prompts de utilidad para el profesorado para utilizar IA generativa* [En línea] [consult. 2025-06-18]. Disponible en: https://openaccess.uoc.edu/bitstream/10609/147885/2/Manual%20IA_PROMTS_CAS.pdf.
- INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2013. *ISO 25964-2:2013, Information and documentation. Part 2: Interoperability with other vocabularies*. Ginebra: ISO.
- INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2011. *ISO 25964-1:2011. Information and documentation. Part 1: Thesauri for information retrieval*. Ginebra: ISO.
- KELLY, Sage, Sherrie-Anne KAYE, y Oscar OVIEDO-TRESPALACIOS, 2023. What factors contribute to the acceptance of artificial intelligence? A systematic review. *Telematics and Informatics* [En línea]. **77**, 1-33 [consult. 2025-06-18]. Disponible en: https://scholar.google.com/scholar?hl=es&as_sdt=0,5&q=What+factors+contribute+to+the+acceptance+of+artificial+intelligence?+A+systematic+review&btnG=.
- MARTÍNEZ DE MADARIAGA, Ricardo, 2023. ¿Bibliotecas inteligentes? Comentarios sobre inteligencia artificial aplicada en bibliotecas. *Revista de la Red de Bibliotecas del CSIC* [En línea]. **39**(91), 91-99 [consult. 2025-06-18]. Disponible en: <https://doi.org/10.20350/digitalCSIC/15390>.
- MILES, Alistair, y Sean BECHHOFFER, 2009. *SKOS Simple Knowledge Organization System Reference. W3C Recommendation* [En línea] [consult. 2025-06-18]. Disponible en: <https://www.w3.org/TR/skos-reference>.
- MOHSEN, Ghassan, et al., 2018. On the automatic construction of an Arabic thesaurus. En: *2018 9th International Conference on Information and Communication Systems (ICICS)*, Nueva Jersey [En línea]. Nueva Jersey: IEEE, pp. 243-247 [consult. 2025-06-18]. Disponible en: <https://ieeexplore.ieee.org/abstract/document/8355431>.
- OPENAI, 2025. *ChatGPT* [En línea] [consult. 2025-06-18]. Disponible en: <https://chat.openai.com/chat>.
- REITZ, Joan M, 2013. *Online Dictionary for Library and Information Science* [En línea] [consult. 2025-06-18]. Disponible en: <https://odlis.abc-clio.com>.
- SANDERSON, Katharine, 2023. GPT-4 is here: what scientists think. *Nature* [En línea]. **615** [consult. 2025-06-18]. Disponible en: <https://doi.org/10.1038/d41586-023-00816-5>.
- SHEIKH, Haroon, Corien PRINS, y Erik SCHRIJVERS, 2023. Artificial intelligence: definition and background. En: *Mission AI: the new system technology* [En línea]. Cham, Suiza: Springer, pp. 15-41 [consult. 2025-06-18]. Disponible en: https://link.springer.com/chapter/10.1007/978-3-031-21448-6_2.
- VOSSSEN, Piek, 2022. ChatGPT is a waste of time. En: *VU Magazine* [En línea] [consult. 2025-06-18]. Disponible en: <https://vumagazine.nl/professor-piek-vossen-chatgpt-is-a-waste-of-time?lang=en>.