Que quede prolijo

Fabricio Gaviña
Licenciado en Bibliotecología y Documentación · Argentina
https://orcid.org/0009-0007-1776-791X
gavinafabricio@gmail.com

Resumen

La IA ya entró a las bibliotecas para acelerar el procesamiento técnico, pero los modelos globales operan por pura probabilidad estadística. Eso hace que las terminologías regionales terminen tapadas por categorías genéricas. A partir de ejemplos patagónicos como veranada o mallín, este ensayo examina el impacto de reemplazar el vocabulario situado por encabezamientos promedio. La tesis central es que la interoperabilidad entre catálogos no requiere borrar las identidades locales, y la propuesta pasa por usar la tecnología para adelantar borradores sin delegarle al algoritmo la decisión final sobre los descriptores locales.

Palabras clave: Inteligencia artificial; Catalogación; Descriptores locales; Tesauros situados; Vocabulario regional; Práctica bibliotecaria.

Abstract

AI is already cataloging library collections to speed up technical processing, but global models rely strictly on statistical probability. As a result, regional terms often get swallowed by generic categories. Drawing on Patagonian examples like veranada or mallín, this essay examines what happens when situated vocabulary gives way to average subject headings. The central argument is that interoperability does not require erasing local identities, and the solution lies in using AI to draft technical metadata without handing over the final say on local descriptors to the algorithm.

Keywords: Artificial intelligence; Cataloging; Local descriptors; Situated thesauri; Regional vocabulary; Library practice.

El problema empieza cuando todo parece correcto

La IA entró a la biblioteca. No golpeó la puerta, no pidió permiso, ya está sentada catalogando como si trabajara ahí de toda la vida. La discusión de si había que dejarla pasar llegó, como casi todo en esta profesión, un poco tarde.

Mientras se debatía en congresos, el backlog seguía creciendo. Ninguna moción aprobada por unanimidad cataloga un solo libro. Así que no. Esto no es un texto para convencer a nadie de "adoptar la IA con responsabilidad" (esa frase que ya suena a capacitación obligatoria de recursos humanos). Ya la adoptamos. La pregunta que importa es otra, bastante menos cómoda, ¿Qué se pierde en el proceso y quién se hace cargo de eso que se pierde?

Porque la IA no llega neutral, aunque se presente con cara de calculadora. Llega entrenada en un español que técnicamente es el mío, o al menos lo entiendo así, pero que en los hechos es de otro lado, una mezcla estadística de variedades del español que no termina de representar a ninguna. Y cuando ese promedio tiene que catalogar algo que pasó acá, en la Patagonia, algo que tiene un nombre preciso acá y ningún nombre en ningún otro lado, el sistema no se traba. No tira error. Elige la categoría más parecida que encontró y sigue. Como si no hubiera pasado nada. Ese es el problema, en realidad. Que no se note. Que quede prolijo, catalogado, aparentemente bien hecho.

En este texto, llamaremos descriptores locales a los términos incorporados deliberadamente para representar denominaciones geográficas, culturales o productivas propias de una comunidad, como complemento de los vocabularios controlados generales. Este artículo no viene a pedir que se frene la máquina. Viene a decir que alguien tiene que quedarse mirando lo que la máquina da por sentado. Y ese alguien, para variar, sigue siendo el bibliotecario.

La naturaleza probabilística de los modelos

La IA no comprende lo que cataloga en el mismo sentido en que lo comprende un profesional. Calcula relaciones, probabilidades y patrones. Sucede que le muestran millones de textos y aprende, a fuerza de repetición, qué palabra suele ir después de cuál otra. Ninguna magia ahí adentro, solo estadística con buena presencia. Y la estadística tiene mayoría. Si los datos con los que se entrenan estos modelos están sobrerrepresentados por determinadas variedades del español, el sistema puede reproducir esa asimetría como si fuera la norma. El de acá, a lo sumo, queda como variante menor, o ruido de fondo de la muestra.

Esto no es solo un problema de representación estadística, aunque se explique fácil con esas palabras. La investigadora Paola Ricaurte viene insistiendo hace años en que estos sistemas no extraen datos en el vacío, los extraen igual que se extrae cualquier otro recurso de la periferia, sin devolver valor a cambio y sin preguntar qué se pierde en el camino. Su planteo sobre la colonialidad del poder a través de los datos puede trasladarse, con las debidas precauciones, al problema de un catálogo bibliotecario, el dato local entra al sistema como

materia prima, sale convertido en categoría genérica, y el que lo produjo no vuelve a verlo reconocido en la estructura del metadato.

Y esto no arrancó con los modelos de lenguaje, conviene aclararlo. El bibliotecólogo argentino Edgardo Civallero ya lo venía señalando antes de que la IA generativa existiera como tal, sistemas como la CDD, la CDU, el LEMB o la LCC nunca fueron neutrales, cargan con las ideologías de quienes los diseñaron, mucho antes de que ningún algoritmo tocara una sola ficha. La diferencia con la IA es la velocidad a la que ese sesgo ahora se reproduce.

Ahora bien, hay una respuesta razonable a todo esto, y conviene decirla antes de que alguien la piense por su cuenta y crea que no la vimos venir. Quien defiende los tesauros globales podría argumentar, con razón, que la interoperabilidad depende justamente de no fragmentar el lenguaje documental en mil variantes locales. Si cada biblioteca cataloga con sus propios términos y nadie los comparte con nadie, se pierde la posibilidad de buscar entre instituciones, de prestar entre bibliotecas, de que un catálogo colectivo tenga sentido. Ese argumento no es menor, y cualquiera que haya trabajado en catalogación colaborativa sabe que tiene peso real. Un tesauro que se fragmenta demasiado deja de ser tesauro y pasa a ser una colección de idiolectos, cada biblioteca hablando su propio dialecto sin que nadie más la entienda.

El problema es que interoperabilidad y borramiento no son la misma cosa, aunque se los confunda todo el tiempo. Los lenguajes documentales estandarizados pueden complementarse con términos locales, notas de alcance o vocabularios propios de la institución, sin que eso implique necesariamente abandonar el encabezamiento general. La técnica para sostener las dos cosas a la vez existe hace rato y cualquier bibliotecario con dos años de oficio la conoce de memoria. Lo que falta no es la herramienta. Es la decisión de usarla, y alguien que se siente a hacerlo cuando la IA ya entregó el borrador y falta la revisión.

Y acá va la pregunta incómoda para quien está leyendo esto desde su escritorio de catalogación, ¿cuándo fue la última vez que agregó un descriptor local porque el estándar no le alcanzaba? Si la respuesta es nunca, no es porque no haya hecho falta. Es porque nadie se lo pidió, y usted tampoco se lo exigió a nadie.

El valor del metadato situado

Una biblioteca no es un depósito neutral de información. Es un espejo de la comunidad que la sostiene. Y un espejo que devuelve una imagen genérica, prolija pero ajena, no sirve de mucho. Sirve para mostrar, no para encontrar.

Porque de eso se trata un catálogo, al final del día. De que alguien pueda entrar y encontrar lo que busca con las palabras que ya tiene en la cabeza. Nadie llega a preguntar por "ganadería

trashumante", llega preguntando por la veranada, por el arreo, por el nombre que le puso su abuelo. Y no es solo el arreo. En la meseta rionegrina un mallín no es un simple humedal ni una pastura cualquiera, es una vega de altura donde se concentra el agua y la vida en medio de la aridez, y toda una economía familiar depende de saber exactamente dónde está el mallín bueno y cuál se secó el año pasado. Un vocabulario general podría terminar remitiéndolo a categorías como "humedales" o "zonas húmedas", y no es una posibilidad remota. Ya pasa en la propia literatura científica que trata el tema, donde el mallín aparece traducido como wetland o wet-meadow, con el nombre original relegado a una aclaración entre comillas. Y ahí corre el riesgo de desaparecer la palabra que cualquier puestero de la zona usa sin pensarlo dos veces.

Ahí aparece la diferencia entre un tesauro estandarizado y uno situado. El primero ordena el mundo desde algún lugar que, casi siempre, no es este. El LEMB, por ejemplo, hace un esfuerzo genuino por pensar en español, pero sigue construido desde una lógica más bien centralizada, pensada para cubrir lo más posible en el menor número de encabezamientos. Eso funciona bien para lo general. Para lo particular, para lo bien nuestro, se queda muy corto. Y cuando se queda corto, alguien tiene que completar lo que falta. Ese alguien no es la IA. Nunca lo fue.

En la propia bibliotecología existe un concepto para esto. Mario Barité lo llama garantía cultural, la idea de que un sistema de organización del conocimiento necesita legitimarse desde la cultura de la comunidad que lo usa, no solo desde la literatura especializada que lo alimenta. Un tesauro puede estar perfectamente construido según sus propias reglas internas y, aun así, no tener ninguna garantía cultural para quien lo consulta desde la meseta.

Este problema tiene nombre también en la sociología. Boaventura de Sousa Santos lo llama sociología de las ausencias, la idea de que hay saberes que no desaparecen porque se los destruya activamente, sino porque el modelo dominante de producción de conocimiento simplemente no los nombra, y lo que no se nombra, para ese modelo, no existe. No es que alguien se siente un día a decidir que la veranada no importa. Es que el sistema nunca la vio lo suficiente como para aprender que existe, y esa diferencia, entre el olvido decidido y el olvido por ausencia estructural, es exactamente lo que hace tan difícil de combatir este problema. No hay ningún villano a quien señalar con el dedo. Hay, nomás, un vacío que se llenó solo, con lo que había más a mano.

Mantener vivos los términos locales, entonces, no es un capricho identitario ni una nostalgia de folclore. Es, lisa y llanamente, una condición para que el acceso a la información sea real y no solo teórico. Un catálogo que cataloga bien pero que nadie puede usar con sus propias

palabras es un catálogo que fracasó en su función más elemental, aunque quede impecable en la estadística de registros procesados.

Esto en Argentina hace tiempo se está discutiendo, y no solo en tono de alarma. La discusión no es solamente sobre el acceso a la tecnología, es sobre quién decide los términos con los que esa tecnología organiza lo que sabe. Y esa discusión, aunque se dé en papers y en paneles de congreso, todavía no bajó lo suficiente hasta el mostrador de la biblioteca popular de El Bolsón, de una escuela rural de Los Menucos, un instituto técnico en Choele Choel o una sede universitaria de Viedma. La biblioteca, como institución, tiene ahí un lugar que no le puede ceder a nadie más.

Ahora, alguien podría decir, con algo de razón, que esto ya se resuelve con las expansiones locales del LEMB o con las notas de alcance que cada institución agrega según su necesidad, y que no hace falta dramatizar tanto. El asunto es que esas expansiones dependen enteramente de un trabajo humano manual, sostenido, revisado, que es exactamente la capa de trabajo que la automatización tiende a saltear cuando nadie la supervisa. No es que el problema sea nuevo. Es que la velocidad con la que ahora se cataloga hace mucho más fácil que esa capa se pierda sin que nadie lo note a tiempo, porque antes catalogar mal costaba tiempo y ahora catalogar mal no cuesta nada.

Así que la pregunta, de nuevo, es para quien está de este lado del mostrador, ¿cuánto de ese acto es ahorro de tiempo legítimo, y cuánto es simplemente dejar que la ausencia se produzca sola?

La sinergia profesional-algoritmo

Ninguna de las partes anteriores pide volver atrás. Nadie está proponiendo catalogar a mano de nuevo, con fichero de cartón y birome, como quien reivindica una tradición que en realidad nadie extraña. La velocidad que aporta la IA es real y es necesaria. Lo que falta decir es qué lugar le queda al humano en el medio de esa velocidad.

Le queda el lugar de siempre, solo que ahora tiene otro nombre. La IA puede hacer el borrador. Puede tirar una propuesta de encabezamiento en un segundo, puede adelantar una buena parte del trabajo repetitivo que antes le comía la tarde entera a cualquier bibliotecario. Que lo haga. Para eso está.

Lo que no puede hacer, y esto no es una limitación técnica que se vaya a resolver con la próxima versión, es saber lo que no le enseñaron. No sabe que acá veranada no es sinónimo de pastoreo. No sabe que un mallín no es un humedal cualquiera, por nombrar algunos ejemplos. No sabe que hay temas que no se pueden pasar por la licuadora de la neutralidad sin

perder lo esencial. Ricaurte hablaría de un dato que entró como materia prima y salió sin su origen. De Sousa Santos diría que se produjo una ausencia más, chiquita, silenciosa, pero ausencia al fin. Las dos lecturas, aunque vengan de tradiciones distintas, terminan diciendo lo mismo con otras palabras. Alguien tiene que quedarse para que eso no pase.

Y conviene imaginar la escena completa, no solo el argumento abstracto. Un chico de una escuela rural de la meseta busca en el catálogo digital algo sobre el trabajo de su abuelo criancero, y no encuentra nada, porque lo que hay está catalogado como "actividades agropecuarias" a secas. Ese chico no va a pensar que el catálogo tiene un sesgo estructural de entrenamiento. Va a pensar que la biblioteca no tiene nada sobre lo suyo. Y ahí, en ese momento exacto, se termina de confirmar todo lo que Ricaurte y de Sousa Santos escribieron desde teorías más grandes. La ausencia bajó del paper al mostrador, y nadie en el medio la frenó.

Y esto no es una escena inventada para que cierre bien el argumento. Catalina Naumis Peña documentó algo parecido en los catálogos mexicanos, donde los encabezamientos sobre pueblos originarios, traducidos directamente de listas construidas en otro país, terminan sin reflejar el enfoque que prevalece en México desde hace décadas. El chico de la meseta y el catálogo mexicano tienen el mismo problema de fondo, solo que uno está documentado en un paper y el otro todavía no se documentó en ningún lado.

Allí reaparece la necesidad del bibliotecario. No porque la tecnología lo necesite por compasión, sino porque es la figura profesional que puede mirar un registro catalogado y preguntarse, honestamente, esto que dice, ¿es lo que somos, o es lo que le quedó más cómodo a la máquina?

La IA ya entró a la biblioteca. Que se quede. Pero que no se siente sola en el escritorio, y que nadie se llame a sorpresa cuando, en un tiempo, alguien pregunte quién estaba mirando mientras tanto.

Referencias

Barité, M. (2011). La garantía cultural como justificación en sistemas de organización del conocimiento: aproximación crítica. *Palabra Clave (La Plata)*, *1*(1), 2–11. https://palabraclave.fahce.unlp.edu.ar/article/download/PCv1n1a02/1867/2732

Civallero, E. (2005). Espinas ocultas: Lenguajes documentales, ideologías negativas y revisiones. *Fuentes del Congreso. Boletín de la Biblioteca y Archivo Histórico del H. Congreso de Bolivia*, *4*(20), 212–213. https://www.aacademica.org/edgardo.civallero/54

Naumis Peña, C. (2022). Desafíos éticos en la aplicación de los lenguajes controlados. En G. A. Torres Vargas (Coord.), *Desafíos en el entorno de la información y la documentación ante las problemáticas sociales actuales* (Vol. 1, pp. 203–218). UNAM, Instituto de Investigaciones Bibliotecológicas y de la Información. https://ru.iibi.unam.mx/jspui/bitstream/IIBI_UNAM/CL131/1/11_uso_etico_informacion_presentacion_catalina_naumis_pena.pdf

Ricaurte, P. (2019). Data epistemologies, the coloniality of power, and resistance. *Television & New Media*, *20*(4), 350–365. https://doi.org/10.1177/1527476419831640

Santos, B. de S. (2002). Para uma sociologia das ausências e uma sociologia das emergências. *Revista Crítica de Ciências Sociais*, *63*, 237–280. https://doi.org/10.4000/rccs.1285