

Del catálogo de fichas al espacio vectorial

Los modelos de lenguaje como infraestructuras de mediación y memoria social



Diego Ferreyra

Director / *Información, cultura y sociedad*

Universidad de Buenos Aires. Facultad de Filosofía y Letras. Departamento de Bibliotecología y Ciencia

de la Información. Universidad de Buenos Aires. Facultad de Filosofía y Letras. Instituto de Investigaciones

Bibliotecológicas (INIBI). Buenos Aires, Argentina / srdiego@uba.ar | <https://orcid.org/0000-0001-6823-5496>

Resumen

Se traza un paralelismo histórico entre la emergencia del catálogo de fichas a finales del siglo XVIII y la irrupción contemporánea de los Grandes Modelos de Lenguaje (LLM), identificándolos como hitos tecnológicos análogos en la organización del conocimiento. Se argumenta que ambos artefactos surgieron como respuestas técnicas y provisionales a desafíos logísticos de escala. A partir de esta analogía, se analizan desafíos posibles para el campo bibliotecológico. Se concluye que estas problemáticas, aunque se presentan con nueva intensidad, actualizan preguntas fundantes de la disciplina sobre el status del conocimiento legítimo, la interpretación y la circulación material de los artefactos culturales.

Palabras clave

*Catalogación
Modelos de lenguaje (Large
Language Models)
Memoria social
Mediación documental*

From the catalog of index cards to the vector space. Language models as infrastructures of mediation and social memory

Abstract

In this study a historical parallel is drawn between the emergence of the card catalog at the end of the 18th century and the contemporary rise of Large Language Models (LLMs), identifying them as analogous technological milestones in the organization of knowledge. It is argued that both artifacts arose as technical and provisional responses to logistical challenges of scale. Based on this analogy, potential challenges for the field of library science are analyzed. It is concluded that these issues, although presented with renewed intensity, bring to the forefront fundamental questions of the discipline regarding the status of legitimate knowledge, interpretation, and the material circulation of cultural artifacts.

Keywords

*Cataloguing
Large Language Models
Social memory
Documentary mediation*

El final del siglo XVIII supuso al menos dos hitos luminosos para nuestro espacio. La promulgación del primer código de catalogación concebido como un componente en la consolidación de derechos ciudadanos por la Asamblea Nacional Constituyente de la Revolución Francesa en 1791 (Krajewski, 2011; Machado y Zafalon, 2022; Library of Congress, 2007) y la creación de un primer catálogo de fichas utilizado como esquema primario de acceso a acervos documentales. En 1780 el crecimiento material de las colecciones de la Imperial and Royal Library of Vienna (antecedente de la Biblioteca Nacional de Austria) planteaba un desafío técnico y metodológico para el mantenimiento de los catálogos impresos. La disolución de la orden jesuita en 1773 y la consiguiente transferencia de fondos librarios a la Biblioteca Real implicaba un desborde logístico para las metodologías y normativas de la organización. Fue así que Gottfried van Swieten –prefecto de la biblioteca entre 1777 y 1803– estableció una serie de procedimientos para confeccionar un *registro temporario* en fichas de papel con la intención de compilarlas luego en un catálogo encuadernado. Esta solución provisional devino finalmente en el primer catálogo de fichas permanente. Desde esta perspectiva, ambos procesos –el código de catalogación y el catálogo de fichas móviles– constituyeron una respuesta técnica y *provisional* a un desafío logístico de origen político.

La emergencia del catálogo de fichas móviles revolucionó todos los órdenes de la actividad social en Occidente. Ofreció un dispositivo que, a partir de una serie de rasgos arbitrarios atribuidos a un recurso, permitía establecer diversos y múltiples esquemas de organización sobre universos ilimitados. El catálogo de fichas brindó las condiciones de posibilidad para proponer y componer universos alternos, diversos y móviles sin alterar la materialidad de los recursos referenciados. Una solución técnica que permitió gestionar y controlar grandes poblaciones de entidades (libros) a través de referencias simbólicas y la abstracción de rasgos descriptivos (títulos, autores, etc). El catálogo de fichas presentaba un sin fin de ventajas en un mundo de multitudes, mercancías, tráfico y movilidad. Una solución mecanizable que permitía deslocalizar saberes eruditos en una serie de tareas formales, un dispositivo modular, escalable, abierto y plástico: con un conjunto relativamente pequeño de reglas ofrecía capacidades para describir, organizar y representar universos de todo tipo con una elasticidad funcional que le permitía adaptarse a necesidades y contextos puntuales y cambiantes.

A partir de su emergencia, esta tecnología se expandió como un paradigma omnipresente durante al menos doscientos años. Desde la burocracia –en cualesquiera de sus órdenes– y hasta la gestión de memorias sociales, todos los sustratos de la actividad humana se fueron materializando en la forma de un fichero y sus cajones. Soldados, ciudadanos, domicilios, barcos, mercancías, profesiones y documentos fueron plasmados y dinamizados por ficheros y catálogos estableciendo una infraestructura técnica compartida entre los procesos de organización de lo social y el registro y mediación del conocimiento humano. Adicionalmente el catálogo de fichas era una verdadera *cadena de montaje* documental multi-usuario y multi-operario que articulaba tareas y sujetos en su mantenimiento, uso y explotación.

“Ahora, todo parece indicar que el libro, en esta forma heredada de la tradición, se encamina hacia su fin.” Con esta reflexión Walter Benjamin (2002) en la década de 1920 consideraba que las fichas –en este caso las fichas eruditas–, eran una de las formas superadoras del libro como tal, un artefacto superador de la *mediación anticuada del libro*. Según Benjamin, las fichas aislaban los *extractos relevantes* de cada libro y facilitaban la comunicación significativa y polémica entre autores e investigadores. También veía en la materialidad mobiliaria del archivo un retorno a la tridimensión artefactual de otros archivos, como los quipus y runas incaicas. Cien años después disponemos de nuevas propuestas basadas en la identificación, abstracción y extracción de rasgos que luego podrán ser recombinados según diferentes esquemas,

propósitos, audiencias o escenarios. Los grandes modelos del lenguaje (LLM, su sigla en inglés) ofrecen la posibilidad de analizar recursos documentales a través de la formalización y asignación de rasgos en tanto componentes atómicos mínimos (tokens) y su sistematización en archivos computables. Desde el presente es posible observar esta coyuntura como un punto más del pasaje comenzado a mediados del siglo pasado. Desde la certidumbre uniforme y categórica de los modelos booleanos a los esquemas de gradualidades, frecuencias y distribuciones. De las condiciones de igualdad booleana a las de similitud vectorial entre entidades.

Podemos entender un LLM como un tipo de artefacto capaz de identificar estructuras a partir de la serialización estadística de los componentes atomizados de un corpus, documento o recurso. Si el catálogo de fichas es capaz de descomponer un documento en algunas dimensiones (autores, títulos, fechas, etc), la representación vectorial de un documento es capaz de establecer múltiples dimensiones por cada palabra (o token) y luego, eventualmente, computar la combinación de las N dimensiones en, nuevamente, múltiples dimensiones. Donde nuestras capacidades para construir sentido ven casos, anécdotas, hilos de palabras, un LLM es como una ganzúa estilística que identifica patrones de legibilidad y estructuras emergentes a partir de la serialización de millones de rasgos en miles de dimensiones. Es, probablemente, una capacidad de análisis lógico que resulta sobrehumana en términos de racionalidad cognitiva, al igual que vivir bajo el agua, volar o viajar a la Luna (D'Andrea y Blanco, 2015).

Al igual que antes los catálogos de fichas, los LLM constituyen una nueva forma de deslocalización y reuso de bienes simbólicos. Son intrínsecamente dependientes del conocimiento social registrado, al igual que los catálogos, no es posible concebirlos o analizarlos como fenómenos externos a las comunidades que los utilizan y producen. Y, al igual que los catálogos una vez más, verificamos su utilidad y pertinencia en procesos que atraviesan un sinnúmero de espacios, tanto en actividades sociales y económicas como en la mediación y producción del conocimiento y la memoria social. Son verdaderos archivos de la intersubjetividad de un momento o de una comunidad, quizás como nunca antes permiten capturar y documentar prácticas discursivas en todo tipo de contextos. Es aquí donde creo que podemos encontrar algunas semillas problemáticas que invocan los arados de nuestro espacio disciplinar para florecer. Quisiera compartir al menos tres aspectos surgidos en estos meses en discusiones e intercambios que tuvieron lugar en diferentes actividades del INIBI.

- » **Ultra-normalización:** la normalización, la adopción consensuada de modelos ideales que organizan actividades e interacciones, fue inicialmente un logro revolucionario y burgués en contra de la monarquía (Canguilhem, 1971). Una vez más, el catálogo, el papel moneda y prácticamente todas las infraestructuras sociales conllevan procesos de estandarización y normalización. Los LLM son archivos, dependen del pasado registrado. La utilización del pasado registrado como insumo y su combinación con dinámicas productivas de escala supra-industrial podría ofrecernos un mundo de bienes simbólicos ultra-normalizados. Nuestra intervención podría favorecer propuestas que permitan la contingencia idiosincrática de modelos locales y variados de sentido.
- » **Usos sociales de los LLMs:** uno de los compromisos del conocimiento científico está basado en la elaboración de premisas reproducibles más allá del sujeto experimentador. En teoría, diferentes experimentadores obtendrán los mismos resultados si se siguen los mismos pasos en las mismas condiciones. Esto implica un desafío importante en relación a los LLMs. Su dinámica o heurística está organizada según modelos estadísticos de probabilidades, de relaciones entre insumos de ingreso (inputs) y resultados (outputs), no es

posible utilizar explicaciones o interpretaciones mecánicas o causales para estos. No es que la dinámica de los LLM resulte inexplicable o inentendible, sino que su proceso no es interpretable en términos mecánicos de causas y efectos. Estas condiciones nos invitan a explorar paradigmas de uso que eviten la mitificación oracular y nos acerquen a la comprensión operativa de sus capacidades y limitaciones.

- » **Los LLM como recurso:** los LLMs son artefactos de memoria y conocimiento social y, como tales, constituyen un potencial de acción e intervención social. Como todo conocimiento socialmente construido, amplía las capacidades de agencia de sujetos, proyectos u organizaciones para legitimar cosmovisiones o llevar adelante las transformaciones que consideren necesarias para vivir mejor. Y, al igual que cualquier otro artefacto, no incluye a todos por igual, ni acceden todos por igual, ni articula a todos los actores por igual. Podemos encontrar aquí una interpelación legítima para construir herramientas genuinas que garanticen la identificación, preservación, acceso y uso de LLMs creados con los recursos, bienes, archivos y memorias de nuestras comunidades.

Estas alternativas y discusiones pueden parecer candentes y urgentes, pero disponemos de un sinnúmero de ejemplos que nos recuerdan que las tensiones y conflictos a partir de cambios en las condiciones materiales de los procesos de mediación y circulación discursiva no son accidentes extraordinarios o únicos. Podemos encontrar en el Fedro (Platón 1983) una de las primeras discusiones registradas al respecto. Platón nos comparte un diálogo entre Sócrates y Fedro donde el primero se opone a dejar por escrito sus enseñanzas. Según Platón, Sócrates identifica tres aspectos que lo invitan a desconfiar en la escritura como medio para la externalización y socialización del conocimiento. Señala que aquellas personas que hubieran accedido a sus enseñanzas a través de la escritura “darán la impresión de conocer muchas cosas, a pesar de ser en su mayoría unos perfectos ignorantes; [...] al haberse convertido, en vez de sabios, en hombres con la presunción de serlo.” En segundo lugar considera que la escritura no es un medio fiable capaz de garantizar la correcta interpretación de sus enseñanzas ya que el texto “cuando es maltratado, o reprobado injustamente, constantemente necesita de la ayuda de su padre, pues por sí solo no es capaz de defenderse ni de socorrerse a sí mismo.” Y por último, plantea su preocupación por las condiciones de diseminación material misma: “.... basta con que algo se haya escrito una sola vez, para que el escrito circule por todas partes lo mismo entre entendidos que entre aquellos a los que no les concierne en absoluto, sin que sepa decir a quiénes les debe interesar y a quiénes no.”

Tenemos pues aquí tres ejes que atraviesan nuestra disciplina y constituyen quizás el fuego que motoriza nuestras prácticas: ¿qué es para cada comunidad o contexto social conocimiento auténtico y legítimo? ¿Cuáles son las herramientas sociales disponibles para facilitar ciertas interpretaciones y obturar otras en los procesos de producción de sentido? y por último, ¿qué dinámicas de circulación material de los artefactos culturales necesitan y reclaman nuestras comunidades? Estas preguntas permanentes e imperfectas encaminan nuestra agenda y tareas de siempre.

Referencias bibliográficas

- » Benjamin, Walter. 2002. *Calle de mano única*. Madrid: Editora Nacional.
- » Canguilhem, George. 1971. *Lo normal y lo patológico*. Buenos Aires: Siglo XXI
- » D'Andrea, Aldana y Javier Blanco. 2015. ¿Pueden nadar los submarinos? La tesis de Turing y el mecanicismo. En Ahumada, José. V., Alfonso Nicolás Venturelli y Silvio Seno Chibeni, eds. *Filosofía e Historia de la Ciencia en el Cono Sur. Selección de trabajos del IX Encuentro y las XXV Jornadas de Epistemología e Historia de la Ciencia*. Córdoba: Asociación de Filosofía e Historia de la Ciencia del Cono Sur. Universidad Nacional de Córdoba. <<http://www.afhic.com/wp-content/uploads/2019/01/pueden-nadar-los-submarinos.pdf>> [Consulta: 30 noviembre 2025].
- » Krajewski, Markus. 2011. *Paper Machines: About Cards & Catalogs, 1548-1929*. Massachusetts: The MIT Press.
- » Library of Congress. 2007. *The card catalog: books, cards, and literary treasures*. San Francisco: Chronicle Books.
- » Machado, Raildo de Souza y Zaira Regina Zafalon. 2020. *Catálogo: dos princípios e teorias ao RDA e IFLA LRM*. João Pessoa: Editora UFPB.
- » Platón. 1983. *El Banquete; Fedón; Fedro*. Barcelona: Orbis.