



Embeddings y modelos conversacionales: una nueva forma de explorar, comprender, describir y conectar la información en bibliotecas

FERNANDO MARTÍNEZ DE GUZMÁN

Ingeniero de datos en Fundación Juan March

Las bibliotecas y centros de documentación se enfrentan a un cambio de paradigma en la recuperación y gestión de información, impulsado por la irrupción de los embeddings o representaciones vectoriales del significado lingüístico del texto. Este artículo explora qué son los embeddings, cómo funcionan y qué nuevas capacidades aportan al ámbito bibliotecario. Se examina su aplicación en tareas como la búsqueda semántica, el enriquecimiento automático de metadatos, la desambiguación de autoridades y la recomendación de documentos. Asimismo, se describe la arquitectura de generación aumentada por recuperación (Retrieval-Augmented Generation, RAG), que combina la búsqueda semántica con modelos de lenguaje conversacionales para construir sistemas capaces de responder preguntas complejas manteniendo la trazabilidad con las fuentes originales. A partir de pruebas de concepto y desarrollos experimentales en curso, el artículo muestra el potencial de los archivos conversacionales como una nueva manera de acceso a las colecciones, especialmente para la exploración transversal y el descubrimiento de relaciones no evidentes. Se analizan también las implicaciones profesionales y éticas de estas tecnologías para las bibliotecas, subrayando que lejos de sustituir la labor bibliotecaria, la refuerzan y amplían. El artículo concluye que la integración crítica y reflexiva de embeddings y RAG puede contribuir a una nueva generación de bibliotecas semánticas e interactivas, alineadas con los valores fundamentales de fiabilidad, trazabilidad y acceso al conocimiento.

Bibliotecas digitales; organización del conocimiento; inteligencia artificial; embeddings; recuperación de información; búsqueda semántica; generación aumentada por recuperación; RAG

Libraries and documentation centers are undergoing a paradigm shift in information retrieval and management, driven by the emergence of embeddings—vector representations of linguistic meaning. This article explores the nature of embeddings, their underlying mechanisms, and the novel capabilities they introduce to the library sciences. It examines their application in tasks such as semantic search, automated metadata enrichment, authority disambiguation, and document recommendation systems. Furthermore, the paper describes the Retrieval-Augmented Generation (RAG) architecture, which integrates semantic search

with conversational language models to build systems capable of addressing complex queries while maintaining traceability to original sources. Drawing on current proofs of concept and ongoing experimental developments, this article demonstrates the potential of "conversational archives" as a new modality for collection access, particularly for cross-sectional exploration and the discovery of non-obvious relationships. The professional and ethical implications of these technologies for libraries are also analyzed, emphasizing that far from replacing library work, they reinforce and expand it. The article concludes that the critical and reflexive integration of embeddings and RAG can contribute to a new generation of semantic and interactive libraries, aligned with the core values of reliability, traceability, and access to knowledge.

Digital libraries; knowledge organization, artificial intelligence, embeddings, information retrieval, semantic search, Retrieval-Augmented Generation, RAG

1. Introducción: hacia un nuevo paradigma en la recuperación de información

Las bibliotecas y los centros de documentación han desarrollado tradicionalmente sistemas para organizar y facilitar el acceso al conocimiento. Clasificaciones, encabezamientos de materia, tesauros y metadatos estructurados constituyen una arquitectura que ha permitido describir de forma rigurosa grandes colecciones de documentos, tanto impresos como digitales. Este entramado conceptual, fruto de un profundo trabajo bibliotecario, ha sostenido la accesibilidad de la información en múltiples disciplinas y contextos (Hjørland, 2016).

Sin embargo, la transformación digital iniciada en las últimas décadas y acelerada en los últimos años ha modificado radicalmente las expectativas de los usuarios. La consulta en lenguaje natural se ha convertido en el patrón de interacción dominante; los sistemas de información han pasado de funcionar como catálogos estancos a convertirse en plataformas de descubrimiento; y la explosión de contenido digital, combinada con la creciente complejidad de los fondos documentales, ha generado un desafío evidente: ¿cómo garantizar el acceso significativo en colecciones que crecen más rápido de lo que pueden ser descritas manualmente? (McCay-Peet et al., 2015; Carevic et al., 2018).

En este contexto ha emergido una tecnología de inteligencia artificial que está transformando este ámbito: los *embeddings*, o representaciones vectoriales del significado lingüístico. Su capacidad para capturar similitudes conceptuales más allá de las palabras permite que los sistemas de información comparen textos por su significado global y no únicamente por coincidencias literales. Este enfoque se inscribe en una evolución más amplia de la recuperación de información hacia modelos semánticos y neuronales que superan los límites de la búsqueda tradicional basada en términos (Manning, Raghavan, & Schütze, 2008; Mitra & Craswell, 2018).

Cuando estas representaciones se integran con modelos de lenguaje conversacionales mediante arquitecturas de generación aumentada por recuperación (Retrieval-Augmented Generation, RAG), se abren nuevas posibilidades para interactuar con los fondos documentales, sintetizar información, navegar por colecciones complejas y generar respuestas fundamentadas exclusivamente en documentos verificables.

Este artículo explica en qué consisten los *embeddings*, por qué representan un cambio estructural en la forma de tratar la información y qué nuevas capacidades aportan a las bibliotecas. Describe también cómo se están realizando pruebas de concepto y desarrollos

iniciales, materializados en entornos de acceso abierto a la comunidad, basados en estas tecnologías, y plantea finalmente qué horizontes se abren para la gestión documental y el apoyo a la investigación.

2. Qué son los embeddings y por qué representan un cambio de paradigma

En el ámbito del procesamiento del lenguaje natural, un *embedding* es una representación matemática que codifica el significado lingüístico de un texto y su contexto, normalmente en forma de un vector de cientos o miles de dimensiones. Cada palabra, frase o documento es transformado en un conjunto de valores numéricos calculados por modelos entrenados sobre grandes corpus textuales. Estas representaciones distribuidas del significado fueron introducidas inicialmente por modelos como word2vec y desarrolladas posteriormente en modelos contextuales como BERT, que tienen en cuenta el contexto completo del texto (Mikolov et al., 2013; Devlin et al., 2019).

Lo relevante no es el valor numérico en sí, sino la estructura del espacio vectorial que generan estas representaciones: textos semánticamente similares ocupan posiciones próximas entre sí, aunque no compartan vocabulario literal. Esta propiedad cambia profundamente la manera en que los sistemas pueden interactuar con el contenido textual. Al representar los documentos mediante vectores, es posible comparar textos por su significado global y no únicamente por las palabras concretas que emplean.

Para las bibliotecas, esto supone una ampliación sustancial de sus capacidades tradicionales. La representación vectorial permite buscar por significado, reduciendo la brecha entre el lenguaje del usuario y el lenguaje de los documentos. Permite también describir por contenido, identificar afinidades temáticas, relacionar documentos por conceptos no explícitos y enriquecer colecciones mediante análisis automático del texto completo. Desde esta perspectiva, los *embeddings* no deben entenderse como un complemento técnico de la búsqueda, sino como una herramienta conceptual que redefine la relación entre el usuario y la colección documental. Constituyen la base de una nueva manera de comprender, describir y conectar la información en bibliotecas.

Este cambio de paradigma no se limita exclusivamente al tratamiento del texto. Los mismos principios de representación vectorial se están aplicando también a otros tipos de contenido, como imágenes, audio o vídeo, dando lugar a *embeddings* visuales, sonoros o multimodales que amplían las posibilidades de análisis, descripción y acceso a las colecciones digitales. Estos enfoques permiten establecer relaciones transversales entre distintos formatos y favorecen una gestión más integrada de fondos heterogéneos, en línea con la diversidad creciente de materiales que conservan y difunden las bibliotecas y los centros de documentación.

3. Más allá de la búsqueda semántica: nuevas capacidades para bibliotecas y centros de documentación

Aunque la búsqueda semántica es la aplicación más visible de los *embeddings*, su potencial va mucho más allá. La posibilidad de representar el significado y el contexto de los textos de forma cuantitativa permite abordar tareas que tradicionalmente han requerido un elevado esfuerzo

manual o una intervención especializada constante, y abre nuevas vías para apoyar los procesos de organización, descripción y acceso a la información (Soergel, 2009).

Una de las aplicaciones más relevantes en este ámbito es el enriquecimiento automático de metadatos. A partir del análisis del contenido completo de los documentos, los modelos pueden sugerir descriptores, identificar temas recurrentes, detectar entidades relevantes o proponer categorías emergentes (Al-Natsheh et al., 2018; Vlachidis et al., 2013). Estas técnicas no sustituyen la labor del profesional de la información, pero sí la complementan, permitiendo orientar el trabajo humano hacia tareas de revisión, curación e interpretación, y contribuyendo a una gestión más ágil de colecciones en crecimiento.

En este marco, algunas de las pruebas de concepto desarrolladas se han centrado en aplicaciones concretas con impacto directo en los procesos bibliotecarios. Entre ellas destaca la desambiguación de términos, especialmente en el ámbito de las autoridades, donde el uso de representaciones vectoriales ha permitido mejorar de forma apreciable tanto los tiempos de cálculo como la precisión en determinados casos, en particular en contextos caracterizados por ambigüedad terminológica, variantes lingüísticas o solapamientos conceptuales.

Asimismo, se han explorado mecanismos de asignación automática de etiquetas y recomendaciones de clasificación, basados en la similitud semántica entre documentos, como apoyo a los procesos de descripción y organización del contenido. Estas experiencias ponen de manifiesto el valor de los *embeddings* como herramienta complementaria a los sistemas tradicionales, orientada a reducir carga operativa y a favorecer una mayor coherencia descriptiva.

Junto a estas aplicaciones ya exploradas, los *embeddings* facilitan otras capacidades con un alto potencial para bibliotecas y centros de documentación, como la agrupación temática emergente, la detección de duplicados o versiones, la mejora del acceso a textos afectados por errores de OCR o la recomendación de documentos relacionados a partir de afinidades conceptuales profundas. Aunque estas posibilidades no han sido aún objeto de pruebas sistemáticas, forman parte del horizonte que abre la representación semántica del contenido y apuntan hacia formas más dinámicas y flexibles de gestión documental.

En conjunto, estas capacidades contribuyen a una concepción renovada de la organización del conocimiento, en la que los sistemas no se limitan a reflejar estructuras predefinidas, sino que son capaces de aprender del contenido, sugerir relaciones y ofrecer nuevas formas de exploración y análisis. Esta dimensión semántica emergente no reemplaza los modelos bibliotecarios tradicionales, pero los complementa, ampliando sus posibilidades en un entorno digital cada vez más complejo.

4. Generación aumentada por recuperación: una arquitectura para el descubrimiento

La convergencia entre *embeddings* y modelos de lenguaje ha dado lugar a una arquitectura que está ganando presencia en bibliotecas y centros de investigación: la generación aumentada por recuperación (Retrieval-Augmented Generation, RAG). Este enfoque combina la recuperación de documentos relevantes mediante búsqueda semántica densa con la generación de respuestas sintéticas basadas exclusivamente en esos documentos (Lewis et al., 2020; Karpukhin et al., 2020).

El funcionamiento del sistema es conceptualmente sencillo. En primer lugar, la consulta del usuario se transforma en un *embedding* que se compara con las representaciones vectoriales de la colección. A continuación, se recuperan fragmentos conceptualmente relevantes, que sirven como contexto para un modelo de lenguaje encargado de sintetizar la información. De este modo, la respuesta generada no procede del conocimiento interno del modelo, sino del contenido documental previamente recuperado.

Este enfoque resulta especialmente relevante en el ámbito bibliotecario, ya que permite construir sistemas conversacionales capaces de responder a preguntas complejas sin perder trazabilidad ni fiabilidad, manteniendo el vínculo con las fuentes originales.

5. Archivos conversacionales: una nueva forma de acceso a las colecciones

La incorporación de modelos de lenguaje conversacionales introduce una modalidad de acceso radicalmente distinta a las colecciones documentales: la conversación con memoria y contexto. A diferencia de los sistemas tradicionales de búsqueda, que tratan cada consulta como una acción aislada, los sistemas conversacionales son capaces de mantener el hilo temático, interpretar reformulaciones, gestionar referencias implícitas y adaptar progresivamente sus respuestas a la intención del usuario.

Este tipo de interacción resulta especialmente relevante en el ámbito bibliotecario, donde las necesidades informativas rara vez se presentan de forma completamente definida desde el inicio. En muchos casos, el usuario parte de una idea o de una pregunta exploratoria, que se va concretando a medida que avanza el proceso de consulta. Los sistemas conversacionales permiten trasladar esta dinámica al entorno digital, ofreciendo una experiencia más cercana a la referencia asistida tradicional.

Desde el punto de vista técnico, esta capacidad se apoya en la combinación de *embeddings* y arquitecturas RAG. Los *embeddings* permiten representar tanto las preguntas como los documentos de forma semántica, mientras que la recuperación aumentada por generación garantiza que las respuestas se construyan a partir de fragmentos documentales concretos. El resultado es un sistema que no solo responde, sino que dialoga, contextualiza y acompaña el proceso de exploración del usuario.



Figura 1. Esquema de un Archivo conversacional basado en generación aumentada por recuperación (RAG).

Las pruebas de concepto y los prototipos funcionales de buscadores conversacionales en los que estamos trabajando muestran el potencial de este enfoque. En lugar de limitarse a devolver listas de resultados, estos sistemas permiten resumir contenidos, comparar enfoques, señalar relaciones conceptuales o sugerir nuevas líneas de exploración basadas en el diálogo previo, de modo que la navegación deja de ser una sucesión de búsquedas independientes para convertirse en un recorrido intelectual guiado por la conversación. Estos desarrollos se están articulando como herramientas experimentales abiertas a la comunidad, con el objetivo de evaluar su utilidad real, la calidad de las respuestas y su integración en los flujos de trabajo bibliotecarios.

Este cambio no implica la sustitución de los catálogos ni de los sistemas de búsqueda existentes, sino su ampliación. Los archivos conversacionales se configuran como una capa adicional de acceso, especialmente adecuada para la exploración inicial, la investigación transversal y el descubrimiento de relaciones no evidentes. En este sentido, representan una evolución natural de los servicios bibliotecarios hacia formas de interacción más flexibles, adaptativas y centradas en el usuario.

6. Implicaciones para bibliotecas y centros de documentación

La adopción de *embeddings* y arquitecturas RAG tiene implicaciones profundas para las bibliotecas y los centros de documentación, que van más allá de la mejora tecnológica puntual. En primer lugar, transforma la relación entre el usuario y la colección. El acceso deja de depender exclusivamente del conocimiento previo del vocabulario controlado, de la estructura del catálogo o de los sistemas de clasificación, y pasa a apoyarse en la expresión natural de necesidades informáticas complejas.

Este cambio resulta especialmente significativo en contextos de investigación interdisciplinar, donde los límites entre disciplinas, terminologías y tradiciones documentales son difusos. La búsqueda semántica y los sistemas conversacionales permiten atravesar estas fronteras

conceptuales, facilitando el acceso a documentos relevantes desde múltiples perspectivas y reduciendo las barreras de entrada a colecciones especializadas.

En segundo lugar, estas tecnologías amplían y redefinen el papel del profesional de la información. Lejos de sustituir la labor bibliotecaria, los sistemas basados en *embeddings* refuerzan la importancia de tareas como la curación de datos, la validación de resultados, la supervisión de los procesos automáticos y la definición de criterios de calidad y relevancia. El bibliotecario se convierte en mediador entre los modelos automáticos y la comunidad usuaria, garantizando que las herramientas semánticas se alineen con los valores y objetivos de la institución.

Además, la introducción de inteligencia artificial en los procesos de recuperación y descripción documental plantea cuestiones éticas y profesionales que deben ser abordadas de forma explícita. La transparencia de los sistemas, la trazabilidad de las respuestas, la gestión de sesgos y la preservación de la diversidad conceptual de las colecciones son aspectos centrales en el diseño y evaluación de estas soluciones. En este sentido, las bibliotecas tienen la oportunidad de situarse como referentes en el uso responsable y crítico de la inteligencia artificial aplicada a la información.

Por último, estas tecnologías abren nuevas posibilidades para la investigación y el análisis documental. La exploración semántica de grandes colecciones permite identificar patrones discursivos, analizar la evolución de conceptos, detectar influencias o conexiones entre documentos y facilitar estudios comparados a gran escala. Estas capacidades refuerzan el papel de las bibliotecas como infraestructuras de conocimiento activas, no solo como repositorios de información.

7. Conclusión: hacia bibliotecas semánticas e interactivas

Los *embeddings* y las arquitecturas de generación aumentada por recuperación representan un cambio profundo en la forma en que las bibliotecas pueden organizar, describir y ofrecer acceso a sus colecciones. No se trata simplemente de una mejora incremental de los sistemas de búsqueda existentes, sino de la incorporación de una nueva capa semántica que permite comprender y relacionar los documentos a partir de su significado.

La integración de estas tecnologías con modelos conversacionales abre la puerta a formas de interacción más naturales, exploratorias y adaptativas, alineadas con la manera en que las personas formulan preguntas, desarrollan intereses y construyen conocimiento. Las pruebas de concepto y desarrollos experimentales actualmente en curso muestran que este enfoque es viable y que puede aportar un valor significativo tanto a usuarios finales como a profesionales de la información.

La apertura de estas herramientas experimentales a la comunidad permite además incorporar la evaluación crítica de los usuarios como parte del propio proceso de desarrollo, reforzando el carácter colaborativo de las bibliotecas como infraestructuras de conocimiento. Al mismo tiempo, la adopción de *embeddings* y RAG no implica abandonar los principios que han definido históricamente a las bibliotecas. Al contrario, estas tecnologías pueden reforzar valores fundamentales como la fiabilidad, la trazabilidad, el acceso equitativo al conocimiento y el rigor en la descripción documental. La clave reside en integrarlas de forma crítica, reflexiva y alineada con la misión institucional.

En este horizonte, el catálogo deja de ser únicamente un instrumento de localización para convertirse en una infraestructura de conocimiento capaz de generar nuevas formas de comprensión, síntesis y descubrimiento. Las bibliotecas que exploren estas tecnologías desde una perspectiva experimental y responsable estarán mejor posicionadas para responder a los desafíos de la era digital, ampliando su papel como espacios de mediación, investigación y construcción colectiva del conocimiento.

Referencias

- Al-Natsheh, H. T., Martinet, L., Muhlenbach, F., Rico, F., & Zighed, D. A. (2018). Metadat enrichment of multi-disciplinary digital library: A semantic-based approach. In *Digital Libraries for Open Knowledge (TPDL 2018)* (Lecture Notes in Computer Science, Vol. 11057, pp. 32–43). Springer. https://doi.org/10.1007/978-3-030-00066-0_3
- Carevic, Z., Lusky, M., van Hoek, W., & Mayr, P. (2018). Investigating exploratory search activities based on the stratagem level in digital libraries. *International Journal on Digital Libraries*, 19(2), 231–251. <https://doi.org/10.1007/s00799-017-0210-1>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 4171–4186.
- Hjørland, B. (2016). Knowledge organization. *Knowledge Organization*, 43(6), 475–484.
- Karpukhin, V., Oguz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D., & Yih, W. (2020). Dense passage retrieval for open-domain question answering. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, 6769–6781.
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge intensive NLP tasks. *Advances in Neural Information Processing Systems*, 33, 9459–9474.
- McCay-Peet, L., Quan-Haase, A., & Kern, D. (2015). Exploratory search in digital libraries: A preliminary examination of the use and role of interface features. *Proceedings of the Association for Information Science and Technology*, 52(1), 1–4. <https://doi.org/10.1002/prai.2015.145052010070>
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of Word representations in vector space. *Proceedings of the International Conference on Learning Representations*.
- Mitra, B., & Craswell, N. (2018). An introduction to neural information retrieval. *Foundations and Trends in Information Retrieval*, 13(1), 1–126.

- Soergel, D. (2009). Digital libraries and knowledge organization. In *Semantic Digital Libraries* (pp. 9–39). Springer. https://doi.org/10.1007/978-3-540-92247-9_2
- Vlachidis, A., Binding, C., May, K., & Tudhope, D. (2013). Automatic metadata generation in an archaeological digital library: Semantic annotation of grey literature. In *Computational Linguistics (Studies in Computational Intelligence, Vol. 458, pp. 187–202)*. Springer. https://doi.org/10.1007/978-3-642-41498-5_10

Sobre el autor



FERNANDO MARTÍNEZ DE GUZMAN

Ingeniero de datos en Fundación Juan March.

Licenciado en Biología por la Universidad Complutense de Madrid. Comenzó a trabajar en el departamento de tecnología de la Fundación Juan March en 2001, realizando labores de administración de sistemas y desarrollo de aplicaciones. En 2017 se incorporó al Datalab de la Biblioteca de la March como ingeniero de datos, haciéndose cargo del ecosistema de herramientas de software libre: sistema de preservación digital, repositorio de difusión, catálogo bibliográfico, sistema de archivo y diversos sistemas analíticos. En los últimos años ha orientado su trabajo hacia la aplicación de técnicas de inteligencia artificial —especialmente embeddings y arquitecturas RAG— al ámbito bibliotecario y documental, explorando nuevas formas de recuperación de información e interacción con las colecciones. Ha participado en diferentes congresos, cursos y seminarios sobre manejo de datos y tecnologías de inteligencia artificial en bibliotecas.