

Strategy for Tacit Knowledge Management in Knowledge Economy SMEs

Federico Francesconi 

Universidad Católica de Santiago del Estero (UCSE), Rafaela, Santa Fe, Argentina
ffedericofrancesconi@gmail.com

Abstract. This paper addresses the challenges of preserving and transferring tacit knowledge (TK) in knowledge economy SMEs (small and medium-sized enterprises), where strategic decisions, skills, intuitions, heuristics, and operational practices often remain scattered within daily interactions and are lost due to staff turnover or the departure of key personnel. A strategy is proposed to capture TK from both social and passive sources, transform it into retrievable representations, and enable access via a conversational agent. The methodology is based on a literature review and comparative analysis, complemented by an exploratory study through a survey administered to 14 industry leaders, experts, and directors. Through this process, different types of TK are identified, classified, and selected; sources and elicitation methods are prioritized; and a collection process is designed. Furthermore, a transformation process is defined, yielding two semantic outputs: text fragments associated with embeddings (semantic vector representations) and RDF triplets (Resource Description Framework) aligned with a domain ontology. The results include the comprehensive documentation of the strategy and a high-level architecture adaptable to the target SMEs. The study concludes that the strategy can improve TK conservation, although due to its nature, it can only be partially formalized, and empirical validation remains as future work.

Keywords: Tacit Knowledge Management, Knowledge Economy SMEs, HybridRAG, Conversational Agent, Knowledge Graphs

Estrategia para la gestión del conocimiento tácito en PyMEs de la economía del conocimiento

Resumen. Este trabajo aborda la problemática de la preservación y transferencia del conocimiento tácito (CT) en pequeñas y medianas empresas (PyMEs) de la economía del conocimiento, donde decisiones estratégicas, habilidades, intuiciones, heurísticas y prácticas operativas quedan dispersas en interacciones cotidianas y se pierden con la rotación o la salida de personas clave. Se propone una estrategia para capturar CT desde fuentes tanto sociales como pasivas, transformarlo en

representaciones recuperables y habilitar su acceso mediante un agente conversacional. La metodología de trabajo se sustentó en una revisión bibliográfica y análisis comparativos, complementados con un estudio exploratorio mediante una encuesta a 14 líderes, expertos y directores del sector. A partir de este proceso se identificaron, clasificaron y seleccionaron diferentes tipos de CT, se priorizaron fuentes y métodos de obtención, y se diseñó un proceso de obtención. Además, se definió un proceso de transformación que produce dos salidas semánticas: fragmentos textuales asociados a embeddings (representaciones vectoriales semánticas) y tripletas RDF (Resource Description Framework) alineadas con una ontología de dominio. Los resultados incluyen la documentación completa de la estrategia y una arquitectura de alto nivel, adaptable a las PyMEs objetivo. Se concluye que la estrategia puede mejorar la conservación del CT, aunque por su naturaleza este solo puede formalizarse parcialmente y la validación empírica permanece como trabajo futuro.

Palabras Clave: Gestión del Conocimiento tácito, PyMEs de la Economía del Conocimiento, HybridRAG, Agente conversacional, Grafos de conocimiento

1. Introducción

El conocimiento en una organización puede existir en dos formas: como conocimiento explícito, que se encuentra documentado y sistematizado en manuales, procedimientos o bases de datos; y como **conocimiento tácito** (en adelante CT), que reside en la experiencia, habilidades e intuiciones de los colaboradores. A diferencia del primero, el CT es poco visible y difícil de comunicar por medios tradicionales («ISO 30401», 2018; Nonaka & Takeuchi, 1995; Polanyi, 1983), y en este trabajo se limitó a aspectos técnicos, funcionales y culturales de las tecnologías de la información y la comunicación.

La gestión del CT plantea un desafío crítico en las pequeñas y medianas empresas del sector de la economía del conocimiento (en adelante PyMEs del conocimiento) (Congreso de la Nación Argentina, 2019). En estas organizaciones, y especialmente en el rubro de software y servicios informáticos, la dificultad para conservar y transferir este saber puede poner en riesgo la continuidad y la calidad de los proyectos (Tiwana, 2004), agravado ante la alta rotación de personal (Avellaneda, 2022). Cuando un colaborador clave deja la organización, se pierde CT que nunca llegó a formalizarse. Esto no solo retrasa proyectos vigentes, sino que incrementa el costo de mantenimiento de productos finalizados, afectando la calidad del producto final (Tiwana, 2004). El CT requiere de años de experiencia para interiorizarse (Nonaka & Takeuchi, 1995). Su gestión también permite disminuir la redundancia de consultas directas y mejorar la autonomía de los equipos. Las estrategias tradicionales de gestión del conocimiento presentan limitaciones para hacer accesible y reutilizable el conocimiento individual (Maulini et al., 2018). Además, factores sociales y culturales como la resistencia a compartir información suponen barreras adicionales (Mohamed, 2008).

Para comprender y justificar el problema en el contexto local, durante julio de 2025 se realizó un estudio exploratorio. Este consistió en una encuesta autoadministrada a 14

líderes, expertos y directores del sector en la ciudad de Rafaela y la región. De estos, 12 afirmaron que reemplazar un colaborador con experiencia es complejo y se pierde CT, y 10 coincidieron en que el CT es “Muy importante” para el funcionamiento de su empresa.

El objetivo de este trabajo fue proponer y documentar una estrategia para obtener, transformar, almacenar y acceder al CT inherente a la labor técnica y funcional de los colaboradores en PyMEs del conocimiento, incluyendo una arquitectura de alto nivel de la solución. Además de estos resultados, se ejemplificaron las posibles herramientas para materializar la propuesta. Estos entregables se formularon de manera genérica, de modo que puedan adaptarse a la mayoría de PyMEs del sector sin requerir cambios drásticos. La implementación y validación empírica se reservan para instancias futuras, habiéndose realizado consultas conceptuales con referentes del sector.

1.1. Objetivos específicos

Para lograr el objetivo general se plantearon los siguientes objetivos específicos.

1. Identificar y describir el CT que se busca conservar.
2. Documentar las fuentes relevantes de donde puede obtenerse dicho conocimiento.
3. Analizar y comparar las diferentes alternativas para obtenerlo.
4. Proponer y documentar un proceso para obtenerlo.
5. Definir el proceso de transformación que deberá recibir el conocimiento obtenido.
6. Proponer y documentar las herramientas para el almacenamiento del conocimiento transformado.
7. Proponer una arquitectura de alto nivel que sirva como guía para implementar la estrategia en un modelo genérico de PyME de la economía del conocimiento.
8. Ejemplificar las herramientas y tecnologías para implementar la arquitectura propuesta.

2. Desarrollo y descripción de las fases de la estrategia

La estrategia para la gestión del CT propuesta se organiza en un conjunto de fases consecutivas y complementarias, y surgen de los objetivos específicos planteados.

2.1. Identificación y clasificación del conocimiento tácito

Se definió el CT a gestionar partiendo de una taxonomía empírica de nueve tipos basada en la fuente del conocimiento (Dinur, 2011). Luego se buscaron ejemplos de CT presentes en 13 trabajos académicos que abordan la temática desde distintos enfoques conceptuales (ver Tabla 1). Se seleccionaron aquellos que pueden existir en las PyMEs del conocimiento, y se unificaron los similares, conservando 6 tipos de CT: **Saber técnico aplicado, Criterios técnicos implícitos, Intuiciones y modelos mentales, Conocimiento distribuido en equipos, Conocimiento contextual del negocio o entorno y Prácticas sociales de trabajo**. Su presencia en la práctica se validó analizando 12 publicaciones de empleo del sector de la economía del conocimiento. Por último, se

Tabla 1. Frecuencia de aparición de los componentes del conocimiento tácito en la literatura académica analizada.

Componente	Número de apariciones
Know-how Técnico	13
Heurísticas y reglas empíricas	9
Decisiones situadas y racionales informales	9
Intuiciones y modelos mentales	8
Conocimiento distribuido en equipos (saber quién sabe qué)	7
Experiencia acumulada y lecciones aprendidas	7
Conocimiento contextual (dominio, entorno, cultura)	6
Rutinas y prácticas informales	5
Conocimiento emocional y relacional	4
Meta-conocimiento (conocimiento sobre el conocimiento)	3
Saber qué (tácito declarativo)	2

clasificaron por dimensión (técnica o cognitiva) y grado de expresividad (Nonaka & Takeuchi, 1995; Polanyi, 1983), mostrando que el CT no es homogéneo, sino diverso, lo que coincide con la estrategia planteada en este trabajo.

En el estudio exploratorio los participantes consideraron como más relevantes y difíciles de transmitir al “Conocimiento contextual del negocio”, el “Saber técnico aplicado” y las “Intuiciones y modelos mentales”, dejando como menos importante a las “Prácticas sociales de trabajo”. El “Conocimiento distribuido en equipos” fue señalado como el tipo de CT más sencillo de difundir y representar.

2.2. Documentación de las fuentes relevantes de conocimiento tácito

Entendiendo por fuente aquello que hace de portador de CT, se encontraron 12 en la literatura analizada (Dreyer & Wynn, 2016; Maulini et al., 2018; Mohamed, 2008; Noor & Rana, 2023; Payró-Campos & Vasconcelos, 2021; Salazar Montes et al., 2018), y se seleccionaron 9: **personas clave, proceso de inducción, reuniones presenciales, videollamadas y medios virtuales sincrónicos, canales de comunicación escrita interna, documentos colaborativos informales, tableros de gestión, artefactos de software embebidos y charlas técnicas internas.**

En la encuesta, la fuente considerada más relevante fue “Personas clave” con 12 votos, seguida por “Reuniones presenciales” y “Charlas técnicas internas”, ambas con 6. Es evidente una preferencia por las fuentes de carácter social, relegando las fuentes pasivas¹ y más susceptibles de una extracción automatizada, como “Artefactos de código”, “Documentos colaborativos”, o “Canales de comunicación interna”, la cual tuvo solo 2 votos.

¹ Con fuentes “pasivas” se hace referencia a aquellas que pueden extraerse sin la necesidad de la participación de un colaborador de forma activa en el proceso.

2.3. Análisis comparativo de alternativas para obtener el conocimiento tácito

Primero se analizó la accesibilidad técnica² de las fuentes identificadas: algunas son altamente digitalizables y otras requieren interacción humana (Mohamed, 2008), justificando la necesidad de **estrategias híbridas**. Ciertos autores definen al CT como un fenómeno social que debe ser estimulado (Dreyer & Wynn, 2016), pero en este trabajo se lo considera un activo que puede capturarse. Las alternativas para obtener el CT encontradas en la literatura académica se compararon en base a cinco criterios: grado de intervención humana requerida, dependencia tecnológica, fuentes de CT que utiliza, nivel de formalización logrado y facilidad de implementación en PyMEs. Representados en la Tabla 2, permiten evaluar la viabilidad práctica y el alcance de cada método.

Tabla 2. Comparativa integral de alternativas para obtener conocimiento tácito basada en criterios de viabilidad para PyMEs.

Alternativa	Intervención Humana	Dep. Tecnológica	Fuentes CT	Formalización	Facilidad PyMEs	Fuente
Entrevistas estructuradas	Muy Alta	Baja	Personas; Reuniones	Medio	Alta	(Devora, 2014)
Cuestionarios multimedia	Alta	Media	Personas clave	Medio	Alta	(Devora, 2014)
Diagramación procesos	Alta	Media	Personas; Reuniones	Alto	Media	(Devora, 2014)
Registro experiencias	Alta	Media	Personas; Documentos	Medio	Media	(García, 2012)
Program slicing	Muy Baja	Muy Alta	Artefactos software	Alto	Baja	(Normantas & Vasilecas, 2013)
Reconocimiento patrones	Baja	Alta	Artefactos; Tableros	Medio	Media	(Normantas & Vasilecas, 2013)
Framework MARBLE	Baja	Muy Alta	Artefactos software	Muy Alto	Baja	(Normantas & Vasilecas, 2013)
Extracción pasiva	Media	Muy Alta	Canales; Documentos; Artefactos	Muy Alto	Media	(Seghroucheni et al., 2025a)
Extracción y chatbot	Media	Muy Alta	Documentos; Artefactos; Personas	Muy Alto	Baja	(Fenoglio et al., 2022)
Juego de roles	Muy Alta	Media	Personas; Inducción; Reuniones	Medio	Baja	(Fenoglio et al., 2022)
Extracción pasiva	Media	Alta	Canales; Documentos; Tableros	Alto	Media	(Noor & Rana, 2023)
Agente conversacional	Baja	Muy Alta	Personas (sim.); Canales	Muy Alto	Baja	(Zuin et al., 2025)
Entrevistas de salida	Muy Alta	Baja	Personas clave	Medio	Alta	(Summerscales, 2024)

El análisis realizado se contrastó con la encuesta a referentes del sector. Contrario al enfoque técnico que se le dio a la solución, las alternativas consideradas más aplicables a las empresas fueron las más sociales. Elegido por 13 de los 14 participantes está la “Mentoría y observación directa”, seguido por “Entrevistas y cuestionarios regulares” con 5, y “Entrevistas de salida” con 4. Las dos alternativas más técnicas fueron las menos votadas; el “Proceso automatizado que extrae el CT de fuentes pasivas”, y el “Chatbot que interactúa con los empleados”, recibieron ambos 3 votos.

2.4. Diseño del proceso de obtención del conocimiento tácito

Para componer el proceso de obtención del CT se seleccionaron 3 de los métodos encontrados en la literatura.

Varias de las alternativas aportan la misma idea: automatizar la obtención del CT desde múltiples fuentes pasivas, procesando la información con NLP (Natural Language Processing). Pese a que se catalogaron como más difíciles de implementar, suponen un mayor impacto y requieren menos intervención humana, por lo que se incluyeron en el proceso como **“extracción de fuentes pasivas”**. Como mejora se sugiere que los mismos colaboradores tengan una vía para indicar dónde está el CT.

² Al hablar de accesibilidad técnica se hace referencia a la facilidad para digitalizar o automatizar la captura del CT.

Para que el personal colabore, el sistema de obtención de CT tiene que reducir la fricción, actuando en momentos de baja carga cognitiva y en sesiones breves (Kernan Freire et al., 2023). Teniendo esto en cuenta, se seleccionó el **agente conversacional**, que cumple un doble propósito y también hace preguntas, pero espera a que el colaborador inicie la interacción. Esta funcionalidad permite la curación del contenido obtenido y respeta la validación e iteración humana necesarias para que el conocimiento extraído sea considerado útil (Celino et al., 2025). Para la adopción también es clave que el agente se adapte al software de comunicación que ya utiliza cada empresa.

Por último se incluyeron las **entrevistas de salida**, pidiendo retroalimentación del colaborador no solo cuando deja la organización sino también cuando finaliza un proyecto o cambia de rol. Pueden ser útiles para obtener la información que no está dispuesto a liberar en su día a día pero sí al finalizar un ciclo. Las preguntas pueden ser generadas manualmente o por el agente conversacional, incentivando un formato único y coherencia en la morfología de las respuestas.

Para que el proceso sea menos tedioso para los colaboradores se recomienda aplicar gamificación (Celino et al., 2025), llevando un ranking transparente de las personas que más aportan CT y resaltando el autor de cada segmento de CT almacenado en el sistema.

La extracción de fuentes pasivas es acompañada por un proceso de transformación con técnicas de NLP para hacer explícito el CT. La propuesta, representada en la Fig. 1, consiste en que ese proceso también reciba las salidas de las entrevistas y el agente conversacional. En contraste con las entrevistas, el agente interactúa bidireccionalmente con el colaborador, ya que recibe información, pero también la brinda. Es la única vía de acceso al repositorio de CT, el cual almacena la salida del proceso de transformación, cerrando así el ciclo. Esta integración extiende enfoques de trabajos previos que aplican los métodos seleccionados de forma aislada.

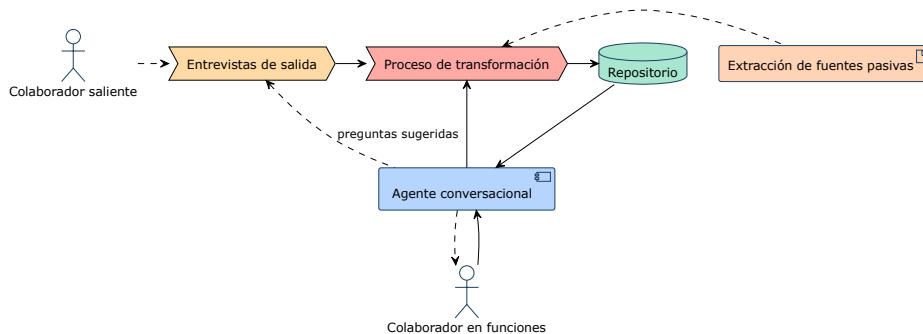


Figura 1. Diagrama de interacción de alto nivel entre los métodos seleccionados para la obtención de CT.

2.5. Definición del proceso de transformación del conocimiento tácito

El primer paso es definir las entradas y salidas semánticas. De las entrevistas de salida se obtiene texto, audio y video, y estos últimos dos pueden ser transcritos a

texto. Al provenir de la oralidad, el contenido se somete a una limpieza estructural y a reconstrucción de la gramática. El agente conversacional recibe respuestas en texto plano, y su pre-procesamiento consiste en filtrar los intercambios funcionales de la conversación que no aportan al mensaje. En la extracción de fuentes pasivas también se hace una conversión a texto, y se recibe el mismo tratamiento que los otros dos métodos. Todo el contenido sin importar el origen se segmenta en fragmentos semánticos con la atomicidad de una oración.

Los metadatos de los fragmentos consisten en un identificador único, el origen (entrevista - agente - fuente pasiva), el identificador del colaborador que lo produce, su rol y área en la organización, y la fecha de obtención del fragmento. Además, se incluye una referencia al contenido original para enriquecer respuestas del agente, y un identificador de tema como una herramienta adicional para agrupar fragmentos relacionados.

Se proponen dos salidas del proceso: **embeddings** (representaciones vectoriales semánticas)³, ya que permiten capturar la semántica de los segmentos, y son robustos frente a la variabilidad lingüística, escalables y eficientes en el acceso; y **tripletas RDF** (Resource Description Framework) alineadas con una ontología, porque tienen lo que los embeddings no: una estructura explícita, capacidad para representar relaciones entre entidades y razonamiento simbólico directo (Celino et al., 2025). Como punto débil, requieren del diseño de una ontología, lo que implica mayores costos. Se eligió RDF frente a lenguajes como OWL (Web Ontology Language) y SKOS (Simple Knowledge Organization System) por su equilibrio entre expresividad, facilidad de implementación y rendimiento en dominios complejos (Seghroucheni et al., 2025a).

Para el diseño del proceso detallado en la Fig. 2 se tuvo en cuenta la literatura académica (Campbell et al., 2025; Seghroucheni et al., 2025b). Los flujos para la generación de las dos salidas se mantienen separados para reducir la dependencia entre sí y facilitar la elección de distintas arquitecturas. Cuando se contemplan ambas, los primeros dos pasos para la generación de las tripletas RDF también pueden utilizarse para enriquecer los embeddings. No se desarrollaron en profundidad las técnicas relativas a cada paso porque otros trabajos ya abordan estos aspectos de forma específica (Ding et al., 2021; Kumar et al., 2020).

Frecuencia de ejecución del proceso de transformación La bibliografía analizada no menciona explícitamente qué frecuencia de ejecución debe tener el procesamiento. Las entrevistas de salida pueden disparar el proceso al cierre de cada una. Para el agente conversacional depende del volumen de fragmentos de CT: si es bajo, dispararlo en cada cierre de conversación; si es alto, agrupar los fragmentos en lotes para ahorrar costos. La extracción de fuentes pasivas, por su alto volumen, se debe procesar en lotes, y de forma inmediata cuando los colaboradores cataloguen el contenido como relevante.

³ <https://www.ibm.com/think/topics/embedding>

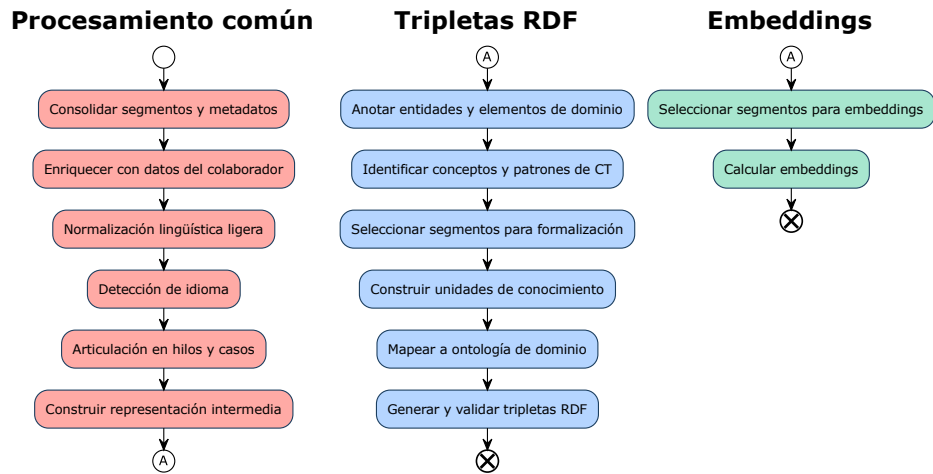


Figura 2. Diagrama de actividades del proceso de transformación del CT.

2.6. Documentación de las tecnologías para el almacenamiento del conocimiento obtenido

Los embeddings y los metadatos mencionados se almacenan en una base de datos vectorial⁴, la cual soporta búsquedas por similitud sobre los vectores. Si bien en los metadatos se incluye la “Referencia al contenido original”, para reducir latencia y garantizar independencia respecto a la mutabilidad de la información también se almacena como campo adicional el texto bruto del segmento.

Para las tripletas RDF y el grafo de conocimiento en el que se integran, el almacenamiento adecuado es una base de datos orientada a grafos compatible con RDF, donde los nodos representan las entidades y las aristas las relaciones entre estas (Hogan et al., 2022). Cada tripleta es un conjunto sujeto-predicado-objeto, y las consultas semánticas se expresan con SPARQL (SPARQL Protocol and RDF Query Language)⁵. Es necesario implementar una ontología específica del dominio de la PyME, a la cual se mapean las tripletas generadas.

2.7. Propuesta de la arquitectura de solución de alto nivel

La arquitectura de alto nivel planteada corresponde a lo que se conoce como RAG (Retrieval-Augmented Generation), técnica en la cual un agente inteligente tiene acceso a un repositorio externo del que puede tomar información para enriquecer sus respuestas. Cuando se utiliza una base de datos vectorial se la denomina VectorRAG y cuando es orientada a grafos GraphRAG. La primera resulta mejor cuando la información está implícita, y la segunda cuando está estructurada explícitamente, siendo superior en fidelidad y precisión de contexto. La combinación de ambas se llama HybridRAG (ver Fig.

⁴ <https://www.ibm.com/es-es/think/topics/vector-database>

⁵ <https://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>

3), y las supera en fidelidad y relevancia de la respuesta (Sarmah et al., 2024). Sin embargo, debido a la mayor complejidad y costo que supone puede no adaptarse a la mayoría de las PyMEs del conocimiento, por lo que solo se recomienda para organizaciones con capacidad económica y técnica elevada. Para el resto se sugiere VectorRAG, una alternativa de menor rendimiento pero más fácil de implementar.

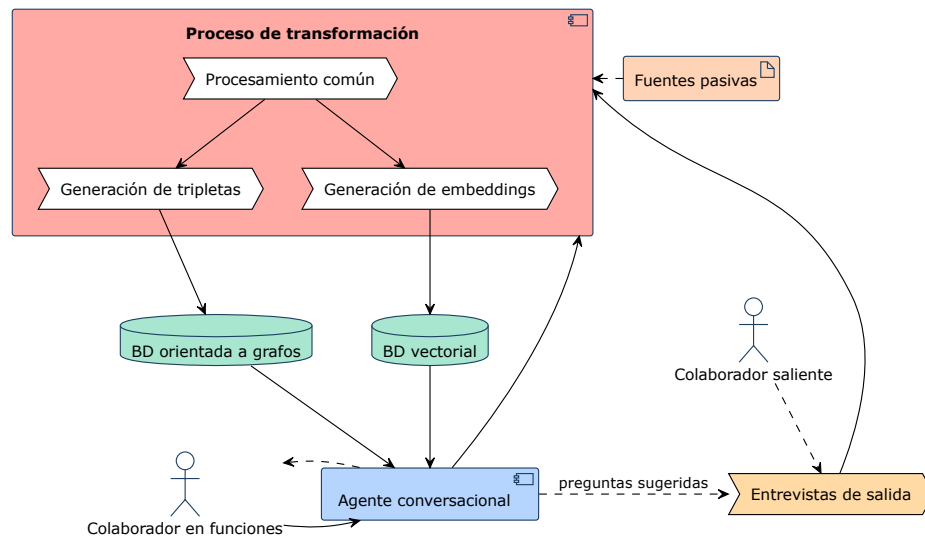


Figura 3. Diagrama de arquitectura de solución basada en HybridRAG. El enfoque VectorRAG prescinde de la base de datos orientada a grafos y la generación de tripletas.

Arquitectura interna y consideraciones sobre el agente conversacional

Detección del conocimiento tácito faltante Siguiendo lo propuesto por la literatura para el CT faltante, el agente cuenta con un estado, acciones, un modelo de transición, y una función recompensa para definir qué acción tomar (Zuin et al., 2025). El estado está compuesto por la estructura de los colaboradores de la organización (que le es dado previamente), y el CT que porta cada uno. En HybridRAG esto se puede plasmar en el grafo de conocimiento, pero en VectorRAG es necesaria una estructura de datos adicional. El objetivo asignado al agente es relevar un porcentaje elevado del CT faltante (definido por cada organización), que se completará progresivamente con sus intervenciones. Después de cada intervención el agente integra el contenido actualizado y lo evalúa críticamente para decidir qué acción tomar. Este razonamiento se apalanca en un LLM (Large Language Model) y es guiado con prompt-chaining.

Privacidad de los datos y anonimización Algunos autores plantean anonimizar la información extraída de las fuentes pasivas antes de analizarla (Noor & Rana, 2023). Sin

embargo, de esa forma se pierde el quién sabe qué, y el agente no puede determinar a quién consultar por el CT faltante. Una posible solución es dejar disponible a los colaboradores una plataforma aislada del proceso de transformación dedicada a información sensible. Otra alternativa es usar identificadores pseudo-anónimos, permitiendo que el agente agrupe segmentos sin revelar la identidad de la persona, pero a costa de no compartir el quién sabe qué con los colaboradores. Por otro lado, pueden existir referencias a contenido original al que el agente o el colaborador no tengan permiso. En este caso la confidencialidad está en el mismo software de la fuente original, que no permitirá el acceso. El agente debe aclarar el inconveniente y generar la respuesta solo con la información ya almacenada.

Arquitectura interna del agente Para reducir la complejidad y aumentar la mantenibilidad, se decidió separar la solución en dos agentes: uno encargado de las consultas de los colaboradores, y otro de la captura y enriquecimiento del CT almacenado. Para tener un razonamiento similar ambos usan el mismo LLM, aunque con instrucciones de sistema distintas y contexto específico a su propósito.

Tanto el formato del campo “Referencia al contenido original” definido en los metadatos, como el protocolo para buscar ese contenido original, dependen de cada plataforma conectada al proceso de transformación. Para desacoplar esta lógica se agrega una “**Capa de herramientas**”, que también incluye las primitivas para acceder a la base de datos vectorial y a la base de datos orientada a grafos, y se comparte por los dos agentes: el de consultas la utiliza para acceder a los repositorios, y el de captura para descubrir el CT faltante y actualizar el existente.

La estructura es modular y se ilustra en la Fig. 4. La primera barrera es un **orquestador** que recibe la entrada del colaborador y decide a qué agente le corresponde. El **gestor de contexto** de la sesión agrega los metadatos del colaborador, permitiendo que el LLM adapte la comunicación. El **módulo de trazabilidad** adjunta a cada respuesta los identificadores de los segmentos obtenidos y las referencias al contenido original. En el caso de HybridRAG, se agrega un **módulo de acceso al conocimiento** para aislar la búsqueda en los dos repositorios y la fusión de los resultados.

2.8. Herramientas y tecnologías para implementar la arquitectura propuesta

A raíz del estudio exploratorio y justificando la arquitectura de alto nivel propuesta, 5 participantes prefirieron un desarrollo propio que les brinde más control, solo 1 se decantó por contratar software externo, y 6 aceptaron un enfoque híbrido. Solamente 3 de los 14 encuestados se mostraron reacios a almacenar el CT de su organización en la nube, lo que avala las herramientas con esta característica.

Base de datos vectorial Este repositorio debe soportar filtros sobre los metadatos asociados a cada segmento e integrarse con el agente conversacional, permitiendo consultas con tiempos de respuesta compatibles con la interacción casi en tiempo real. Para un despliegue en la nube con baja carga operativa para la PyME se sugiere Pinecone⁶. Como alternativa de código abierto con posibilidad de gestión en infraestructura propia existe

⁶ <https://docs.pinecone.io/guides/index-data/indexing-overview>

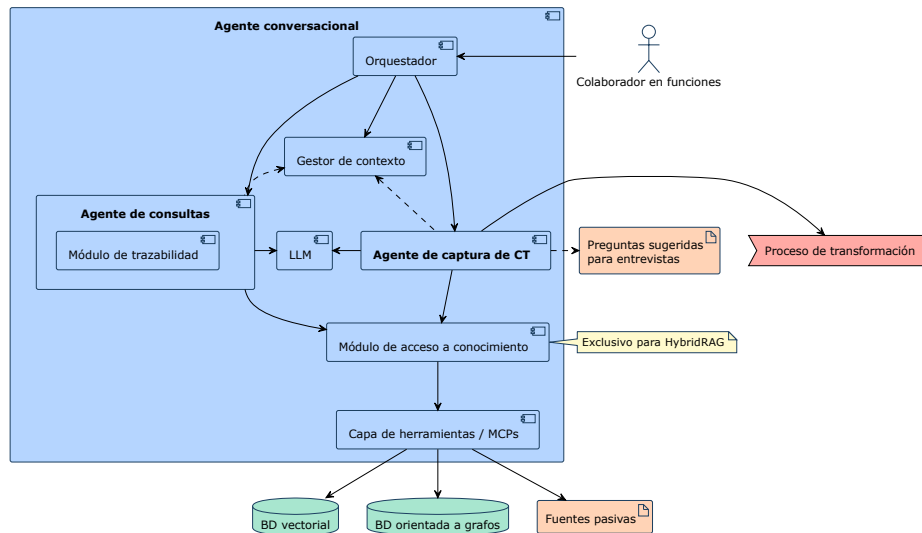


Figura 4. Arquitectura interna del agente conversacional propuesto.

Qdrant⁷. Para aquellas organizaciones que ya utilicen PostgreSQL, una opción atractiva es la extensión pgvector⁸.

Base de datos orientada a grafos El objetivo es soporte para RDF con un costo y complejidad razonables para una PyME. Como herramienta empaquetada y con baja carga técnica se sugiere GraphDB⁹. Una alternativa de código abierto más ligera y auto-hosteable es Apache Jena TDB con Apache Jena Fuseki¹⁰. Otra opción de código abierto equilibrada es Eclipse RDF4J¹¹. Para las organizaciones que prioricen flexibilidad y no quieran usar estándares semánticos como RDF, existen las bases de datos del tipo grafos de propiedades. Un ejemplo es Neo4j¹², con despliegue en infraestructura propia o en la nube, opciones gratuitas o de pago, y que puede incluso integrar una base de datos vectorial interna, permitiendo reducir ambos tipos de repositorios a una única herramienta (Athale & Vaddina, 2025).

Agente conversacional Los marcos de trabajo de código abierto de orquestación de agentes como LangGraph¹³ y Haystack¹⁴ ofrecen componentes reutilizables y patrones para estructurar la lógica. Una alternativa de pago es LlamaIndex¹⁵, un marco de trabajo

⁷ <https://qdrant.tech/documentation/concepts/filtering/>

⁸ <https://github.com/pgvector/pgvector>

⁹ <https://graphdb.ontotext.com/documentation/11.2/>

¹⁰ <https://jena.apache.org/index.html>

¹¹ <https://projects.eclipse.org/projects/technology.rdf4j>

¹² <https://neo4j.com/docs/>

¹³ <https://docs.langchain.com/oss/python/langgraph/overview>

¹⁴ <https://haystack.deepset.ai/>

¹⁵ <https://developers.llamaindex.ai/>

enfocado en RAG que reduce gran parte de la complejidad técnica. No se descarta una implementación propia sin marco de trabajo, ganando control y menos librerías innecesarias a costa de mayor esfuerzo de desarrollo (Campbell et al., 2025).

3. Conclusiones

En línea con los objetivos planteados para el trabajo, se ha definido una estrategia para gestionar el conocimiento tácito en PyMEs de la economía del conocimiento, junto con una arquitectura de alto nivel que sustenta una futura implementación. La mejora frente al estado del arte no se plantea como un nuevo método aislado, sino como una estrategia que combina métodos sociales y técnicos con una arquitectura concreta, adaptable a organizaciones con recursos limitados y necesidades de conservación y transferencia de conocimiento. Aun así, tal como señalan distintos autores en la temática, no es posible obtener la totalidad del CT que reside en los colaboradores, por lo que una parte del mismo permanecerá inefable.

La principal línea de trabajo futuro sería implementar y validar la estrategia propuesta. Como complemento, se destaca la posibilidad de incorporar el procesamiento de otros formatos además del texto plano, lo cual no se aborda aquí debido a la complejidad adicional que introduce. También resulta relevante definir un esquema de evaluación que permita medir el impacto real en contextos organizacionales.

4. Uso de herramientas de inteligencia artificial

Durante la preparación de este trabajo, el autor utilizó herramientas de inteligencia artificial. Para búsquedas dentro de la bibliografía y mejorar la redacción se usaron ChatGPT Plus¹⁶ y Google Gemini Pro¹⁷, y para búsqueda de material bibliográfico se eligió Perplexity¹⁸. Se tuvieron en cuenta propuestas sobre el uso de inteligencia artificial en artículos académicos (Volder & Authier, 2025; Weaver, 2024). Tras utilizar estas herramientas, el autor revisó y editó el contenido según fuera necesario y asume plena responsabilidad por el contenido de la publicación.

Referencias

- Athale, M., & Vaddina, V. (2025). Knowledge Graph Based Repository-Level Code Generation. *2025 IEEE/ACM International Workshop on Large Language Models for Code (LLM4Code)*, 169-176. <https://doi.org/10.1109/LLM4Code66737.2025.00026>
- Avellaneda, P. (2022, noviembre). *Modelo de gestión del conocimiento para un departamento de desarrollo de sistemas en Buenos Aires* [Trabajo Final de Posgrado]. Universidad de Buenos Aires. Facultad de Ciencias Económicas.
- Campbell, C., Boscoe, B., & Do, T. (2025, 25 de julio). *AquiLLM: a RAG Tool for Capturing Tacit Knowledge in Research Groups*. <https://doi.org/10.48550/arXiv.2508.05648>

¹⁶ <https://chatgpt.com/>

¹⁷ <https://gemini.google.com/>

¹⁸ <https://www.perplexity.ai/>

- Celino, I., Scrocca, M., & Chiatti, A. (2025, 17 de marzo). Mutual Understanding between People and Systems via Neurosymbolic AI and Knowledge Graphs. <https://doi.org/10.3233/FAIA250241>
- Congreso de la Nación Argentina. (2019, junio). Ley 27.506 - Régimen de Promoción de la Economía del Conocimiento. <https://www.argentina.gob.ar/normativa/nacional/ley-27506-324101/actualizacion>
- Devora, J. G. U. (2014). *Desarrollo de una herramienta para dar soporte a la extracción del conocimiento tácito en las Pymes de desarrollo de software* [Tesis de Maestría]. Centro de Investigación en Matemáticas (CIMAT), Zacatecas, México.
- Ding, N., Xu, G., Chen, Y., Wang, X., Han, X., Xie, P., Zheng, H.-T., & Liu, Z. (2021, 1 de septiembre). Few-NERD: A Few-Shot Named Entity Recognition Dataset. <https://doi.org/10.48550/arXiv.2105.07464>
- Dinur, A. (2011). Tacit knowledge taxonomy and transfer: Case-based research. *Journal of Behavioral and Applied Management*, 12(3), 246-281. <https://doi.org/10.21818/001c.17865>
- Dreyer, H., & Wynn, M. (2016). Tacit and explicit knowledge in software development. projects: A combined model for analysis.
- Fenoglio, E., Kazim, E., Latapie, H., & Koshiyama, A. (2022). Tacit knowledge elicitation process for industry 4.0. *Discover Artificial Intelligence*, 2(1), 6. <https://doi.org/10.1007/s44163-022-00020-w>
- García, H. (2012). ITAKA: gestión interactiva del conocimiento en organizaciones de desarrollo de software.
- Hogan, A., Blomqvist, E., Cochez, M., D'amato, C., Melo, G. D., Gutierrez, C., Kirrane, S., Gayo, J. E. L., Navigli, R., Neumaier, S., Ngomo, A.-C. N., Polleres, A., Rashid, S. M., Rula, A., Schmelzeisen, L., Sequeda, J., Staab, S., & Zimmermann, A. (2022). Knowledge graphs. *ACM Computing Surveys*, 54(4), 1-37. <https://doi.org/10.1145/3447772>
- ISO 30401:2018. (2018, noviembre). Consultado el 14 de junio de 2025, desde <https://www.iso.org/standard/68683.html>
- Kernan Freire, S., Wang, C., Ruiz-Arenas, S., & Niforatos, E. (2023). Tacit Knowledge Elicitation for Shop-floor Workers with an Intelligent Assistant. *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, 1-7. <https://doi.org/10.1145/3544549.3585755>
- Kumar, A., Pandey, A., Gadia, R., & Mishra, M. (2020). Building Knowledge Graph using Pre-trained Language Model for Learning Entity-aware Relationships. *2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON)*, 310-315. <https://doi.org/10.1109/GUCON48875.2020.9231227>
- Maulini, A., Straccia, L. N., & Pollo Cattaneo, M. F. (2018). Una aproximación a un modelo de gestión de conocimiento aplicable a las pequeñas y medianas fábricas de software [Section: Desarrollo e innovación en ingeniería]. *Desarrollo e innovación en ingeniería, 2018, ISBN 978-958-59127-9-3, págs. 56-63*, 56-63. Consultado el 7 de junio de 2025, desde <https://dialnet.unirioja.es/servlet/articulo?codigo=8717384>
- Mohamed, A. H. (2008). Capturing software-engineering tacit knowledge, 132-137. <https://lens.org/087-087-712-988-004>
- Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company: How japanese companies create the dynamics of innovation*. Oxford University Press.
- Noor, M., & Rana, Z. A. (2023). Natural Language Processing (NLP) based Extraction of Tacit Knowledge from Written Communication during Software Development. *2023 4th International Conference on Advancements in Computational Sciences (ICACS)*, 1-5. <https://doi.org/10.1109/ICACS55311.2023.10089779>
- Normantas, K., & Vasilecas, O. (2013). A systematic review of methods for business knowledge extraction from existing software systems.

- Payró-Campos, P., & Vasconcelos, F. I. F. (2021). Gestión de conocimiento en una empresa de desarrollo de software [Number: 36]. *Revista de Investigación Académica Sin Frontera: Facultad Interdisciplinaria de Ciencias Económicas Administrativas - Departamento de Ciencias Económico Administrativas-Campus Navojoa*, (36). <https://doi.org/10.46589/rdiasf.vi36.422>
- Polanyi, M. (1983). *The Tacit Dimension* (Reprinted). Peter Smith.
- Salazar Montes, D. P., Cárdenas Gaviria, L. A., & Franco, O. H. (2018). Gestión del conocimiento en procesos de desarrollo de software: un marco de trabajo para apoyar a las MiPyMEs. *Scientia et Technica*, 23(1), 76-83. Consultado el 7 de junio de 2025, desde <https://dialnet.unirioja.es/servlet/articulo?codigo=6413735>
- Sarmah, B., Hall, B., Rao, R., Patel, S., Pasquali, S., & Mehta, D. (2024, 9 de agosto). HybridRAG: Integrating Knowledge Graphs and Vector Retrieval Augmented Generation for Efficient Information Extraction. <https://doi.org/10.48550/arXiv.2408.04948>
- Seghroucheni, O. Z., Lazaar, M., & Al Achhab, M. (2025a). Systematic review and framework for AI-driven tacit knowledge conversion methods and machine learning algorithms for ontology-based chatbots in e-learning platforms [Number: 01]. *International Journal of Interactive Mobile Technologies (iJIM)*, 19(1), 126-139. <https://doi.org/10.3991/ijim.v19i01.51051>
- Seghroucheni, O. Z., Lazaar, M., & Al Achhab, M. (2025b). Using AI and NLP for tacit knowledge conversion in knowledge management systems: A comparative analysis [Number: 2]. *Technologies*, 13(2), 87. <https://doi.org/10.3390/technologies13020087>
- Summerscales, J. (2024). Harvesting tacit knowledge for composites workforce development. *Composites Part A: Applied Science and Manufacturing*, 185, 108357. <https://doi.org/10.1016/j.compositesa.2024.108357>
- Tiwana, A. (2004). An empirical study of the effect of knowledge integration on software development performance. *Information and Software Technology*, 46(13), 899-906. <https://doi.org/10.1016/j.infsof.2004.03.006>
- Volder, C. D., & Authier, C. (2025, julio). Presentación: Decálogo para el uso ético de la inteligencia artificial en revistas científicas y académicas [Place: Buenos Aires Published: CAICYT - CONICET]. <https://www.caicyt-conicet.gov.ar/>
- Weaver, K. D. (2024). The artificial intelligence disclosure (AID) framework: An introduction. *College & Research Libraries News*, 85(10), 407. <https://doi.org/10.5860/crln.85.10.407>
- Zuin, G., Mastelini, S., Loures, T., & Veloso, A. (2025, 4 de julio). Leveraging Large Language Models for Tacit Knowledge Discovery in Organizational Contexts. <https://doi.org/10.48550/arXiv.2507.03811>