RESUMEN DEL TRABAJO FIN DE GRADO

Universidad de Salamanca

Escuela Politécnica Superior de Zamora

Título: Cerebro Digital: Aplicación de notas basada en Markdown con integración RAG y análisis semántico topológico.

Autor: Oswaldo Enrique González Suárez

Tutores: María Luisa Pérez Delgado (USAL), Nicolás García Martín (HP SCDS), Adrián Fernández Roldan (HP SCDS)

Titulación: Grado en Ingeniería Informática en Sistemas de Información

Fecha: Julio de 2026

Resumen

La gestión del conocimiento personal (PKM) se enfrenta actualmente al desafío de integrar capacidades analíticas avanzadas de Inteligencia Artificial sin comprometer la privacidad y soberanía de los datos del usuario. El presente Trabajo Fin de Grado describe el diseño, desarrollo y despliegue de *Cerebro Digital*, una plataforma PKM basada en archivos Markdown que incorpora procesamiento de lenguaje natural de forma nativa y estrictamente local. Desarrollado bajo la tutela tecnológica de la empresa HP SCDS, el sistema implementa una arquitectura de microservicios altamente eficiente. El *backend*, construido en Go (Golang) aplicando principios de Arquitectura Limpia, gestiona la concurrencia y la persistencia documental/vectorial en MongoDB; mientras que el *frontend* en React proporciona una experiencia de usuario interactiva y fluida bajo el paradigma de diseño *Glassmorphism*.

Una de las aportaciones técnicas más relevantes del proyecto es su motor de Generación Aumentada por Recuperación (Graph-RAG híbrido). Al orquestar modelos de lenguaje grandes (LLMs) fuertemente cuantizados, como *Qwen3 (0.6b)*, a través del motor Ollama, el sistema es capaz de analizar semánticamente los apuntes del usuario, recuperar el contexto mediante el cálculo algebraico de similitud del coseno y generar respuestas precisas sin requerir conexión a servicios en la nube. Esta decisión arquitectónica erradica los riesgos de brechas de seguridad y garantiza el cumplimiento del RGPD.

Adicionalmente, el aplicativo emplea algoritmos de Teoría de Grafos y simulación de físicas (método de Barnes-Hut) para procesar las referencias cruzadas y los enlaces bidireccionales de los textos, renderizando la información como una red topológica interactiva. Todo el entorno de ejecución ha sido empaquetado bajo estrictos estándares DevOps, utilizando Nix y Docker para asegurar una reproducibilidad determinista. Los *benchmarks* técnicos demuestran que es posible ejecutar flujos RAG complejos y extraer topologías documentales procesando miles de nodos por segundo en hardware de consumo estándar.

Palabras clave: Gestión del Conocimiento Personal (PKM), Inteligencia Artificial Local (Edge AI), Generación Aumentada por Recuperación (RAG), Teoría de Grafos, Arquitectura de Microservicios.

Abstract

Personal Knowledge Management (PKM) systems currently face the challenge of integrating advanced Artificial Intelligence analytical capabilities without compromising user data privacy and sovereignty. This capstone project details the design, development, and deployment of *Cerebro Digital*, a Markdown-based PKM platform that incorporates Natural Language Processing natively and strictly locally. Developed under the technological mentorship of HP SCDS, the system leverages a highly efficient microservices architecture. The backend, built in Go (Golang) applying Clean Architecture principles, manages concurrency and document/vector persistence in MongoDB; meanwhile, the React frontend provides an interactive user experience following the *Glassmorphism* design paradigm.

A core technical contribution of this project is its hybrid Graph-Retrieval-Augmented Generation (Graph-RAG) engine. By orchestrating highly quantized large language models (LLMs), such as *Qwen3 (0.6b)*, through the Ollama engine, the system is able to semantically analyze user notes, retrieve context through algebraic cosine similarity calculations, and generate accurate responses without relying on cloud services. This architectural decision eliminates the risk of security breaches and ensures full GDPR compliance.

Furthermore, the application utilizes Graph Theory algorithms and physics simulations (Barnes-Hut method) to process cross-references and bidirectional links within the texts, rendering the information as an interactive topological network. The entire runtime environment has been packaged under strict DevOps standards, using Nix and Docker to ensure deterministic reproducibility. Technical benchmarks demonstrate the viability of executing complex RAG pipelines and extracting document topologies by processing thousands of nodes per second on standard consumer hardware.

Keywords: Personal Knowledge Management (PKM), Edge AI, Retrieval-Augmented Generation (RAG), Graph Theory, Microservices Architecture.