

SmartNote: Superação das Limitações das Ferramentas de Gestão de Conhecimento Pessoal mediante Ligações Semânticas Automáticas e Inteligência Artificial Local

Daniel Correia-Gonçalves¹, Leonel Morgado^{1,2}

¹ Universidade Aberta, Lisboa, Portugal, 2000065@estudante.uab.pt

² INESC TEC, Porto, Portugal, Leonel.Morgado@uab.pt

Resumo

As ferramentas de gestão de conhecimento pessoal (*Personal Knowledge Management - PKM*), como o Obsidian ou Notion, revelam-se úteis na organização de notas e informações, contudo, apresentam limitações significativas que dificultam a descoberta de relações implícitas, a expansão das capacidades cognitivas e a preservação da soberania dos dados. A sua dependência quase exclusiva de ligações explícitas e manuais tende a gerar fragmentação do conhecimento e sobrecarga informativa, especialmente à medida que o volume de dados cresce. Acresce ainda a vulnerabilidade decorrente da utilização de serviços baseados na *cloud*, que pode comprometer a privacidade e reduz o controlo efetivo do utilizador sobre a sua própria informação. Neste artigo, apresentamos um sistema de gestão de conhecimento pessoal, concebido para mitigar estas limitações através da integração de técnicas de processamento de linguagem natural, indexação vetorial e inteligência artificial local. A solução desenvolvida, designada SmartNote, recorre a *embeddings* semânticos para identificar conexões contextuais, utiliza a biblioteca FAISS para uma indexação vetorial eficiente e integra um assistente de IA baseado em *Retrieval-Augmented Generation* (RAG), totalmente executado em ambiente offline através do Ollama. Os resultados obtidos confirmam a viabilidade técnica desta abordagem e apontam para um novo caminho nos sistemas pessoais de gestão do conhecimento, com maior capacidade de contextualização, escalabilidade e privacidade, proporcionando ao utilizador controlo sobre a sua própria informação.

Palavras-chave: gestão de conhecimento pessoal, ligações semânticas, IA local, privacidade, *embeddings*, FAISS, RAG.

Title: SmartNote: Overcoming the Limitations of Personal Knowledge Management Tools through Automatic Semantic Links and Local Artificial Intelligence

Abstract: Personal Knowledge Management (PKM) tools, such as Obsidian or Notion, prove useful for organising notes and information; however, they present significant limitations that hinder the discovery of implicit relationships, the expansion of cognitive capabilities, and the preservation of data sovereignty. Their near-exclusive reliance on explicit, manually created links tends to generate knowledge fragmentation and information overload, particularly as the volume of data increases. Furthermore, the use of cloud-based

services introduces vulnerabilities that may compromise privacy and reduce the user's effective control over their own information. In this paper, we present a personal knowledge management system designed to mitigate these limitations through the integration of natural language processing techniques, vector indexing, and local artificial intelligence. The developed solution, named SmartNote, employs semantic embeddings to identify contextual connections, uses the FAISS library for efficient vector indexing, and integrates a Retrieval-Augmented Generation (RAG) AI assistant, fully operational offline via Ollama. The results confirm the technical feasibility of this approach and point towards a new direction for personal knowledge management systems, offering enhanced contextualisation, scalability, and privacy, while providing users with control over their own information.

Keywords: personal knowledge management, semantic links, local AI, privacy, embeddings, FAISS, RAG

1. Introdução

A era digital, marcada por um fluxo contínuo de informação, coloca desafios cada vez maiores à capacidade humana de organizar, processar e, sobretudo, de extrair significado do vasto volume de dados disponível [Eppler & Mengis, 2004, Whelan & Teigland, 2010]. A sobrecarga informativa e a consequente fragmentação do conhecimento tornaram-se fenómenos generalizados, afetando tanto profissionais como estudantes que dependem de uma gestão eficaz de notas, documentos e ideias para manter a produtividade e apoiar a aprendizagem contínua [Eppler & Mengis, 2004, Whelan & Teigland, 2010], [Fathizargaran & Cranefield, 2017, Da Silva & Ulbricht, 2024].

Neste cenário, emergiram diversas ferramentas de gestão do conhecimento pessoal (PKM, *Personal Knowledge Management*), como o Obsidian¹, o Notion², o Roam Research³ e o Logseq⁴, concebidas para auxiliar na organização e interligação de informações digitais. Apesar de demonstrarem utilidade na estruturação de dados e na criação de ligações explícitas, a sua arquitetura e as suas funcionalidades revelam limitações estruturais que comprometem a eficácia a longo prazo e dificultam a adaptação ao crescimento das necessidades dos utilizadores. [Clemente & Pollara 2005, Fathizargaran & Cranefield, 2017, Da Silva & Ulbricht, 2024, Thaul, Bleimann & Clarke, 2012, Eppler & Mengis, 2004].

As principais limitações das ferramentas de PKM atualmente disponíveis incluem dificuldades na descoberta implícita de relações, desafios de escalabilidade cognitiva e questões de privacidade e disponibilidade dos dados, fatores que podem reduzir a sua utilidade prática e limitar a autonomia do utilizador, como se detalha na Secção 2.

Este artigo apresenta uma abordagem exploratória para mitigar as três limitações identificadas, através da conceção e implementação do SmartNote, um sistema de gestão de

¹ <https://obsidian.md/>

² <https://www.notion.com/>

³ <https://roamresearch.com/>

⁴ <https://logseq.com/>

conhecimento pessoal que integra técnicas de processamento de linguagem natural (PLN) [Edress & Ortakci, 2024], indexação vetorial [Ghadekar et al., 2023] e inteligência artificial local [Gargari & Habibi, 2025]. O SmartNote indica a viabilidade de um sistema que não se limita a organizar o conhecimento, mas que procura acrescentar valor de forma ativa, ao descobrir relações semânticas e ao disponibilizar um assistente baseado em IA. Tudo isto é alcançado mantendo, em simultâneo, a privacidade e reforçando o controlo do utilizador sobre os seus dados [Gao & Brink, 2019, Baraku et al., 2025], graças a um processamento integralmente local.

Os objetivos específicos deste artigo são:

- Apresentar a arquitetura conceptual do SmartNote, descrevendo de que forma os seus módulos interagem para disponibilizar funcionalidades avançadas de gestão de conhecimento.
- Demonstrar a viabilidade da abordagem proposta através de uma implementação de referência, validando a capacidade do sistema para gerar ligações semânticas automáticas e fornecer assistência inteligente em ambiente totalmente local.
- Discutir as limitações identificadas ao longo do desenvolvimento do SmartNote e as lições daí decorrentes, oferecendo contributos relevantes para futuras investigações e implementações no domínio da gestão de conhecimento pessoal.

O restante artigo encontra-se organizado da seguinte forma: a Secção 2 apresenta os trabalhos relacionados e aprofunda a análise das limitações das ferramentas PKM existentes; a Secção 3 introduz o conceito proposto e descreve a arquitetura do SmartNote; a Secção 4 detalha a implementação e a avaliação do sistema; a Secção 5 discute os resultados obtidos e as suas implicações; por fim, a Secção 6 apresenta as conclusões e sugere direções para trabalho futuro.

2. Trabalhos Relacionados

As ferramentas de PKM, como as já referidas Obsidian, Notion, Roam Research e Logseq, têm transformado significativamente a forma como os utilizadores organizam e interligam notas e ideias. Contudo, apesar dos avanços alcançados, tanto a literatura recente como a experiência prática evidenciam um conjunto de limitações estruturais que impedem estas soluções de explorar plenamente o seu potencial na gestão inteligente do conhecimento [Fathizargaran & Cranfield, 2017, Da Silva & Ulbricht, 2024, Thaul, Bleimann & Clarke, 2012]. Estas limitações podem ser sintetizadas em três categorias centrais, que o SmartNote procura abordar de forma inovadora: descoberta implícita de relações, escalabilidade cognitiva e soberania e disponibilidade dos dados, que constituem limitações estruturais nas ferramentas PKM atuais, justificando a proposta e desenvolvimento do SmartNote como uma alternativa que ambiciona ser mais inteligente, privada e eficaz [Fathizargaran & Cranfield, 2017, Wright, 2005, Wang, Gao & Wang, 2010].

2.1. Descoberta Implícita de Relações

As ferramentas PKM atualmente disponíveis dependem fortemente de ligações explícitas, criadas manualmente pelo utilizador (como a sintaxe `[[links]]` no Obsidian) ou baseadas

em correspondências literais de texto [Wright, 2005]. Embora eficazes para a organização básica de notas, estas abordagens revelam-se insuficientes para identificar e explorar as relações implícitas e semânticas existentes entre diferentes fragmentos de conhecimento [Ackerman & McDonald, 2000, Xu et al., 2024, Da Silva & Ulbricht, 2024].

Por exemplo, uma nota que aborda “*aprendizagem profunda*” pode apresentar uma relação semântica significativa com outra nota sobre “*redes neurais convolucionais*”, mesmo que não exista uma ligação direta ou termos coincidentes. A incapacidade de detetar estas conexões latentes contribui para a fragmentação do conhecimento, mantendo informações relacionadas isoladas em silos digitais e dificultando uma compreensão integrada e a síntese de novas ideias.

Estudos na área da descoberta de informações baseada na literatura técnico-científica (LBD, *Literature-Based Discovery*) e de grafos de conhecimento heterogéneos destacam que a dependência de ligações explícitas e de relações binárias simples pode comprometer a capacidade de inferência de conhecimento. As representações mais sofisticadas, que incorporam relações aninhadas e hipergrafos, são fundamentais para aumentar a precisão e a riqueza do conhecimento inferido [Pu, Beck & Verspoor, 2025, Da Silva & Ulbricht, 2024].

Adicionalmente, *frameworks* como a HIM (*Heterogeneous Information Mining*) evidenciam o potencial de integrar atributos dos nós com informações da estrutura da rede, permitindo a descoberta de relações implícitas em redes sociais heterogéneas, um princípio que se pode transpor para o contexto da gestão de conhecimento pessoal [Xu et al., 2024].

2.2. Escalabilidade Cognitiva

A escalabilidade cognitiva refere-se à capacidade de um sistema permanecer funcional e compreensível à medida que o volume de informação e a complexidade das tarefas aumentam. As ferramentas PKM atuais, embora eficazes para volumes moderados de notas, podem impor uma carga cognitiva excessiva quando o utilizador acumula grandes quantidades de dados: processos como a navegação, a recuperação e a interpretação do conhecimento armazenado tornam-se tarefas exigentes, transformando aquilo que deveria ser um sistema de apoio numa fonte de frustração [Karr-Wisniewski & Lu, 2010, Eppler & Mengis, 2004].

A falta de mecanismos eficientes para sumarizar, contextualizar e apresentar informações de forma inteligente compromete a capacidade do utilizador de manter uma visão clara e acessível do seu próprio conhecimento, limitando a escalabilidade cognitiva do sistema. A utilização de linguagem natural e de resumos linguísticos pode mitigar este desafio, proporcionando uma interface compreensível a utilizadores de diferentes formações [Edress & Ortakci, 2024]. Adicionalmente, a integração de sistemas baseados em IA, capazes de interagir de forma ativa com dados provenientes de múltiplos domínios da vida do utilizador, permite criar uma visão de conhecimento dinâmica e personalizada, potencialmente transformando a gestão de conhecimento pessoal num verdadeiro “colega de IA pessoal” [Aal & Rüller, 2025].

2.3. Soberania e Disponibilidade dos Dados

A soberania dos dados constitui uma preocupação central nas ferramentas PKM, sobretudo no que se refere ao controlo sobre informação pessoal. Muitas das funcionalidades avançadas das ferramentas PKM contemporâneas dependem de serviços baseados na *cloud*. Embora esta abordagem ofereça conveniência na colaboração e acesso ubíquo, ela suscita questões relacionadas com a privacidade, segurança e soberania dos dados do utilizador [Gao & Brink, 2019, Baraku et al., 2025]. Informações pessoais e sensíveis são frequentemente armazenadas em servidores de terceiros, sujeitas às políticas de privacidade dessas entidades e às legislações de jurisdições externas.

No contexto europeu, a transferência de dados pessoais para fora da União Europeia é fortemente regulada pelo Regulamento Geral de Proteção de Dados (RGPD).⁵ Os dados pessoais não podem ser enviados para prestadores de serviços fora da UE sem autorização explícita e, em muitos casos, a transferência para fora da UE não é legítima mesmo com consentimento. Outros países têm vindo igualmente a implementar legislação similar de proteção de dados pessoais, o que reforça a necessidade de soluções PKM que garantam a soberania, a privacidade e o controlo integral sobre os dados dos utilizadores.

Para assegurar a soberania dos dados, é fundamental implementar mecanismos eficazes de governação e controlo de acesso em redes de partilha de dados entre organizações. O conceito de soberania de dados abrange dimensões de proteção, participação e provisão, incluindo controlo, segurança e mecanismos de conformidade destinados a preservar direitos fundamentais durante a partilha de informação. Adicionalmente, *frameworks* como o *Data Capsule* capacitam os indivíduos a exercer soberania sobre os seus dados pessoais, permitindo-lhes definir os termos e condições de utilização [Baraku et al., 2025]. A dependência da conectividade à Internet pode ainda comprometer a disponibilidade da informação em contextos offline, restringindo o controlo e a autonomia do utilizador sobre o seu próprio conhecimento.

Solução	Links manuais	Links automáticos	Semântica	Privacidade local
Obsidian	✓	✗	✗	✓
Roam Research	✓	✓	✗	✗
Notion	✓	✗	✗	✗
Logseq	✓	✓	✗	✓

3. Conceito Proposto

O conceito central deste trabalho consiste num sistema de gestão de conhecimento pessoal capaz de responder diretamente às três limitações identificadas nas ferramentas PKM existentes: a dificuldade na descoberta de relações implícitas, os desafios relacionados com

⁵ <https://eur-lex.europa.eu/eli/reg/2016/679/oj>

a escalabilidade cognitiva e as preocupações relativas à soberania e disponibilidade dos dados. A proposta apresentada assenta na integração sinérgica de tecnologias avançadas de Processamento de Linguagem Natural (PLN) [Edress & Ortakci, 2024], indexação vetorial [Ghadekar et al., 2023] e Inteligência Artificial local [Gargari & Habibi, 2025], garantindo, em simultâneo, privacidade e soberania dos dados [Gao & Brink, 2019, Baraku et al., 2025], e seguindo os seguintes princípios orientadores:

1. *Descoberta de Relações Implícitas*: Para ultrapassar a dependência de ligações explícitas, o sistema proposto recorre à utilização de *embeddings* semânticos e pesquisa vetorial [Edress & Ortakci, 2024], [Ghadekar et al., 2023]. Esta abordagem permite identificar conexões contextuais e temáticas entre diferentes fragmentos de conhecimento, mesmo quando expressos com terminologias distintas ou em documentos separados. Ao representar o significado do texto em espaços vetoriais de alta dimensão, o sistema consegue calcular a similaridade semântica entre notas, revelando relações que permaneciam invisíveis a métodos baseados em correspondência literal. Desta forma, a gestão do conhecimento deixa de ser apenas uma tarefa de organização manual, transformando-se num processo de descoberta assistida por IA, enriquecendo a rede de conhecimento do utilizador [Da Silva & Ulbricht, 2024].
2. *Escalabilidade Cognitiva*: Para atenuar a sobrecarga cognitiva decorrente do aumento do volume de informação, o conceito integra modelos de linguagem locais num fluxo de *Retrieval-Augmented Generation* (RAG) [Gargari & Habibi, 2025]. Este assistente inteligente permite aos utilizadores realizar consultas em linguagem natural sobre o seu conjunto de conhecimento, obtendo respostas contextualmente relevantes e sumarizadas [Edress & Ortakci, 2024, Aal & Rüller, 2025]. Em vez de percorrer manualmente inúmeras notas, o utilizador pode interagir com o sistema de forma conversacional, reduzindo a carga mental e facilitando a compreensão e a síntese de informações complexas. A capacidade de gerar resumos e de responder a questões específicas com base no conteúdo das notas aumenta significativamente a acessibilidade e a usabilidade do conhecimento armazenado [Ghadekar et al., 2023].
3. *Soberania e Disponibilidade dos Dados*: Um dos pilares centrais do SmartNote é a garantia de processamento totalmente local [Gao & Brink, 2019, Baraku et al., 2025]. Ao eliminar a dependência de serviços externos baseados na *cloud*, o sistema assegura privacidade total, controlo e acesso contínuo aos dados do utilizador, independentemente da disponibilidade de ligação à internet. Esta opção de design não só responde às preocupações relacionadas com a soberania dos dados, como também garante que o utilizador mantém a propriedade e o controlo integral da sua informação pessoal, um fator cada vez mais valorizado num contexto digital em que a privacidade é uma preocupação crescente.

A implementação prática deste conceito é materializada no protótipo SmartNote, cuja descrição detalhada será apresentada na secção seguinte. O SmartNote constitui-se como uma prova de conceito, evidenciando tanto a viabilidade técnica como os benefícios práticos desta abordagem integrada para a gestão de conhecimento pessoal.

4. Solução: o SmartNote

O SmartNote foi concebido e desenvolvido como um protótipo funcional que integra diversas tecnologias, permitindo explorar uma abordagem de gestão de conhecimento pessoal com foco em privacidade, escalabilidade e contextualização [Edress & Ortakci, 2024, Ghadekar et al., 2023, Gargari & Habibi, 2025]. A arquitetura modular possibilita a combinação de diferentes componentes, cada um desempenhando um papel específico na pipeline de processamento do conhecimento. O sistema funciona de forma inteiramente local, garantindo a privacidade e reduzindo dependências externas, uma decisão de design central para reforçar o controlo do utilizador sobre os seus dados. [Baraku et al., 2025].

4.1. Arquitetura do Sistema

A arquitetura do SmartNote encontra-se organizada em camadas distintas, de modo a facilitar a manutenção, a extensibilidade e a compreensão do fluxo de dados [Edress & Ortakci, 2024, Ghadekar et al., 2023, Gargari & Habibi, 2025]. A Figura 1 apresenta a arquitetura funcional do sistema, evidenciando os principais módulos e as suas interações.

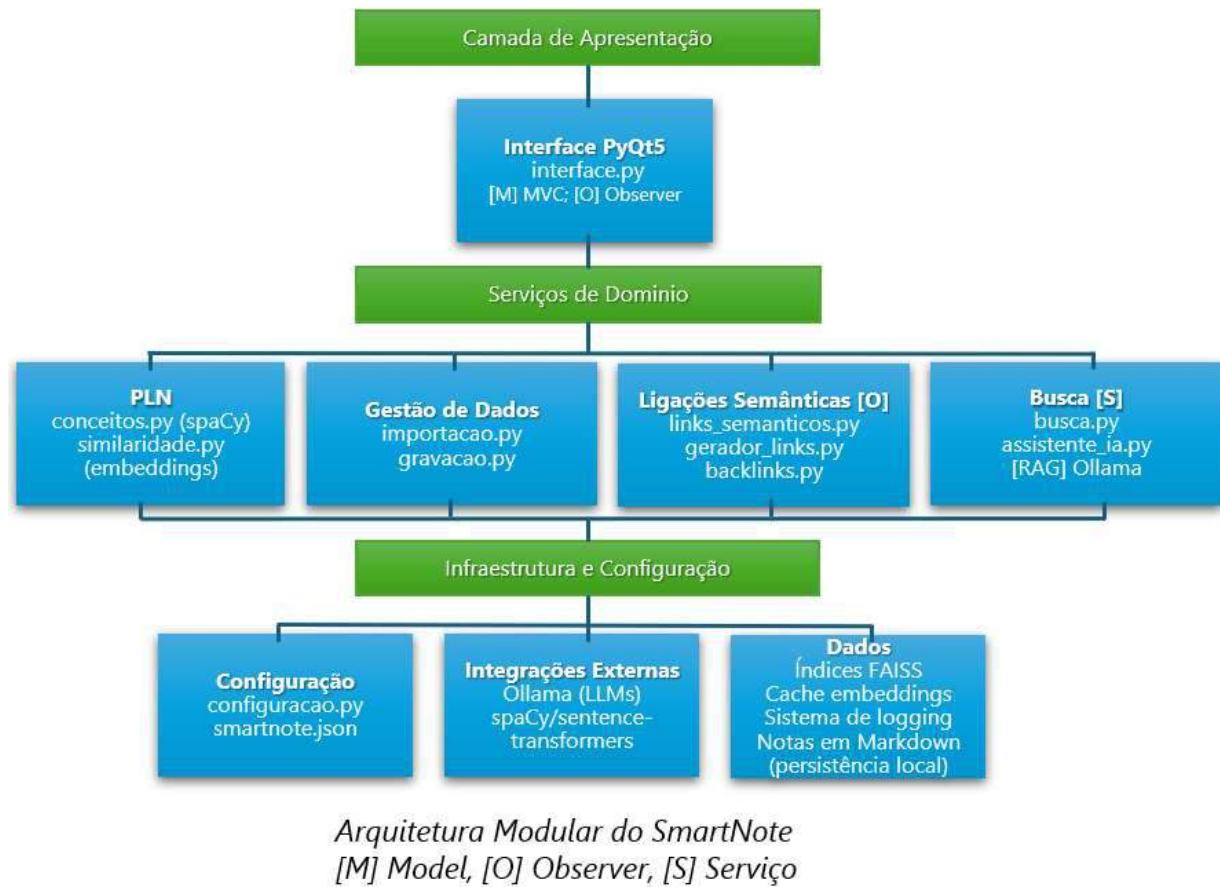


Figura 1. Arquitetura Funcional do SmartNote [M] Model, [O] Observer, [S] Serviço.

Os principais módulos do SmartNote e as tecnologias subjacentes são os seguintes:

- *Módulo de Processamento de Linguagem Natural (PLN)*: Utiliza a biblioteca spaCy para a extração de conceitos relevantes das notas [Alles, Giozza & de Oliveira Alburquerque, 2018]. O ExtratorConceitos identifica entidades nomeadas, substantivos compostos e termos técnicos, indo além da simples correspondência de palavras-chave. Esta escolha baseia-se na robustez e eficiência do spaCy para tarefas de PLN de nível industrial, garantindo uma análise linguística precisa e rápida.
- *Módulo de Representação Semântica*: Recorre à biblioteca sentence-transformers para gerar embeddings semânticos a partir do texto das notas. O modelo paraphrase-multilingual-MiniLM-L12-v2 foi selecionado pela sua capacidade de transformar frases e parágrafos em vetores densos que capturam o significado contextual, permitindo a comparação semântica de textos, e pela sua eficiência, sendo leve e multilíngue. Esta escolha é fundamental para a funcionalidade de descoberta de relações implícitas [Afzal et al., 2025].
- *Módulo de Indexação e Busca*: Utiliza a biblioteca FAISS (Facebook AI Similarity Search) para a indexação e pesquisa eficiente de embeddings vetoriais. O IndexFlatL2 foi escolhido para garantir máxima precisão na busca semântica, recorrendo à distância euclidiana (L2) para medir a similaridade entre vetores. Apesar de poder ser mais intensivo em memória para volumes muito elevados de dados, para o contexto de gestão de conhecimento pessoal com volumes moderados de notas, a sua precisão é preferível a métodos aproximados [Ghadekar et al., 2023].
- *Módulo de Geração de Links e Backlinks*: O GeradorLinksSemanticos identifica e cria ligações entre notas com base na similaridade semântica calculada pelo FAISS. Um limiar de similaridade configurável (atualmente 0,10) determina quando uma ligação deve ser sugerida. O GestorBacklinks assegura a coerência bidirecional das ligações, atualizando automaticamente as notas referenciadas para incluir os respetivos backlinks, criando uma rede de conhecimento interconectada e navegável.
- *Módulo de Assistente IA*: Integra o Ollama para a execução de grandes modelos de linguagem (LLM) localmente, sem necessidade de serviços em cloud. O assistente utiliza uma arquitetura do tipo Retrieval-Augmented Generation (RAG), em que as consultas do utilizador são enriquecidas com informações relevantes recuperadas do corpo de notas indexado. Isto permite fornecer respostas contextualmente precisas e informadas, baseadas no conhecimento pessoal do utilizador, garantindo a privacidade e o controlo dos dados [Gargari & Habibi, 2025].
- *Camada de Interface*: Desenvolvida em PyQt5, a interface gráfica nativa oferece uma experiência que se pretende intuitiva e responsiva. A opção por uma interface nativa, em vez de baseada na tecnologia web, visa assegurar melhor desempenho e integração com o sistema operativo, crucial para um sistema que opera 100% localmente e realiza processamento intensivo de dados.

- *Sistema de Cache Multicamada:* Para otimizar o desempenho e reduzir o reprocessamento desnecessário, o SmartNote incorpora um sistema de cache multicamada. Este armazena resultados de extração de conceitos, *embeddings* já gerados e resultados de pesquisa, acelerando significativamente as operações subsequentes e melhorando a responsividade global do sistema.

4.2. Fluxo de Trabalho

O fluxo de trabalho no SmartNote, em particular no que diz respeito à geração de ligações semânticas, segue uma sequência lógica de etapas, concebidas para assegurar a consistência e a eficiência do processamento [Alles, Giozza & de Oliveira Alburquerque, 2018, Afzal et al., 2025, Ghadekar et al., 2023]:

1. *Importação e Pré-processamento de Notas:* As notas, habitualmente em formato *markdown*, são importadas para o sistema. Nesta fase, é realizado um pré-processamento básico para normalizar o texto e prepará-lo para a análise de PLN.
2. *Extração de Conceitos:* O componente *ExtratorConceitos* recorre ao *spaCy* para identificar termos-chave e conceitos relevantes em cada nota. Estes conceitos são essenciais para a representação semântica do conteúdo.
3. *Geração de Embeddings:* Os conceitos extraídos, e em alguns casos o texto completo das notas, são processados pelo *GeradorEmbeddings* para criar representações vetoriais densas (*embeddings*) utilizando *sentence-transformers*. Estes vetores capturam o significado semântico do texto.
4. *Indexação e Busca por Similaridade:* Os *embeddings* gerados são indexados no FAISS. Quando uma nova nota é adicionada ou uma nota existente é modificada, o sistema realiza uma busca por similaridade no índice FAISS para encontrar outras notas semanticamente relacionadas.
5. *Aplicação de Links e Atualização de Backlinks:* Com base nos resultados da busca por similaridade e num limiar configurável, o *GeradorLinksSemanticos* insere automaticamente ligações ([[conceito]]) no texto das notas. Em simultâneo, o *GestorBacklinks* atualiza as notas referenciadas para incluir os respectivos *backlinks*, garantindo uma navegação bidirecional e uma rede de conhecimento coesa.

Este fluxo de trabalho, aliado à arquitetura modular e às escolhas técnicas implementadas, permite ao SmartNote oferecer uma solução consistente e funcional para os desafios da gestão de conhecimento pessoal, facilitando a descoberta de relações implícitas, a escalabilidade cognitiva e a soberania dos dados.

💡 Fluxo de Geração de Links Semânticos

1. Importação e processamento de notas



2. Extração de conceitos via spaCy



3. Geração de embeddings via sentence-transformers



4. Indexação e busca por similaridade via FAISS



5. Aplicação de links e atualização de backlinks

⚙️ Componentes Críticos

ExtratorConceitos

GeradorLinksSemanticos

GestorBacklinks

AssistenteIAAvancado

Figura 2. Fluxo de trabalho para geração de links semânticos e estado dos componentes críticos no SmartNote.

5. Avaliação e Resultados

A validação do SmartNote foi realizada através de uma metodologia de testes abrangente, com o objetivo de avaliar a robustez, a funcionalidade e a eficiência do sistema, bem como identificar eventuais limitações e extrair lições relevantes para desenvolvimentos futuros. O principal foco desta avaliação consistiu em demonstrar a viabilidade da abordagem proposta para mitigar as limitações das ferramentas PKM existentes, conforme descrito na introdução.

5.1. Metodologia de Validação

Para validar o funcionamento do SmartNote, foi adotada uma abordagem combinada que incluiu testes automatizados e avaliações práticas, seguindo três eixos principais: Componentes, Integração e Desempenho.

Testes de Componentes verificaram o correto funcionamento de partes específicas do sistema, sem validar a eficácia global do conceito. Por exemplo:

O módulo de extração de conceitos foi testado com frases como “A Inteligência Artificial é uma área da ciência da computação...” e “O Machine Learning é uma subárea da IA que

permite aos sistemas aprenderem automaticamente a partir de dados”. Os testes demonstraram que o módulo identificou corretamente os termos mais relevantes presentes em cada frase.

O cálculo de similaridade semântica foi avaliado em cenários controlados, assegurando que conceitos relacionados recebiam pontuações mais elevadas. Como parte dos testes de componentes, foram desenvolvidos testes unitários automatizados utilizando o framework pytest. A Figura 3 apresenta um exemplo da execução desses testes, cobrindo funcionalidades como busca semântica, extração de conceitos, configuração e cálculo de similaridade. Todos os testes foram concluídos com sucesso, reforçando a estabilidade dos módulos individuais.

```
PS C:\SmartNote> python main.py --testes unitarios
>>
2025-06-25 06:24:57,086 - __main__ - INFO - Executando testes do tipo: unitarios
2025-06-25 06:24:57,087 - __main__ - INFO - Executando testes 'unitarios' em: C:\SmartNote\tests\unitarios
===== test session starts =====
platform win32 -- Python 3.10.11, pytest-8.3.5, pluggy-1.5.0 -- C:\Users\operador\AppData\Local\Microsoft\Windows\Terminal\pyt...
cachedir: .pytest_cache
rootdir: C:\SmartNote
plugins: anyio-4.9.0, langsmith-0.3.19
collecting ... collected 25 items

tests/unitarios/test_busca.py::test_busca_simples PASSED [ 4%]
tests/unitarios/test_busca.py::test_busca_ignorar_acentos PASSED [ 8%]
tests/unitarios/test_busca.py::test_busca_diferenciar_maiusculas PASSED [ 12%]
tests/unitarios/test_busca.py::test_busca_regex PASSED [ 16%]
tests/unitarios/test_busca.py::test_busca_sem_resultados PASSED [ 20%]
tests/unitarios/test_conceitos.py::test_extrair_conceitos_basicos_simples PASSED [ 24%]
tests/unitarios/test_conceitos.py::test_extrair_conceitos_basicos_stopwords PASSED [ 28%]
tests/unitarios/test_conceitos.py::test_validar_conceito PASSED [ 32%]
tests/unitarios/test_configuracao.py::test_obter_valor_existente PASSED [ 36%]
tests/unitarios/test_configuracao.py::test_definir_e_obter_valor_temporario PASSED [ 40%]
tests/unitarios/test_configuracao.py::test_exportar_e_importar_configuracoes PASSED [ 44%]
tests/unitarios/test_configuracao.py::test_resetar_configuracoes PASSED [ 48%]
tests/unitarios/test_gravacao.py::test_gravar_nota_simples PASSED [ 52%]
tests/unitarios/test_gravacao.py::test_gravar_varias_notas PASSED [ 56%]
tests/unitarios/test_importacao.py::test_importar_arquivo_com_frontmatter PASSED [ 60%]
tests/unitarios/test_importacao.py::test_importar_arquivo_sem_frontmatter PASSED [ 64%]
tests/unitarios/test_importacao.py::test_importar_diretorio_vazio PASSED [ 68%]
tests/unitarios/test_importacao.py::test_importar_arquivo_com_codificacao_invalida PASSED [ 72%]
tests/unitarios/test_similaridade.py::test_similaridade_identica PASSED [ 76%]
tests/unitarios/test_similaridade.py::test_similaridade_ortogonal PASSED [ 80%]
tests/unitarios/test_similaridade.py::test_similaridade_parcial PASSED [ 84%]
tests/unitarios/test_similaridade.py::test_similaridade_com_vetores_2d PASSED [ 88%]
tests/unitarios/test_similaridade.py::test_entrada_unidimensional_e_bidimensional PASSED [ 92%]
tests/unitarios/test_similaridade.py::test_similaridade_vetores_normalizados PASSED [ 96%]
tests/unitarios/test_similaridade.py::test_similaridade_vetores_aleatorios PASSED [100%]

===== 25 passed in 13.90s =====
2025-06-25 06:25:16,725 - __main__ - INFO - Testes concluídos com sucesso.
PS C:\SmartNote> []
```

Figura 3. Execução de testes unitários no SmartNote.

Nos testes de integração, foram realizados quatro fluxos completos de utilização, incluindo importação de notas, extração automática de conceitos, geração de relações implícitas e acesso ao assistente de IA local. Nos testes realizados, o SmartNote conseguiu gerar ligações semânticas entre conteúdos, mesmo na ausência de correspondência literal entre termos, sugerindo que o sistema se mantém responsável em cenários habituais de utilização pessoal.

Por fim, os testes de desempenho avaliaram a escalabilidade em diferentes volumes de dados (100 notas e 1000 parágrafos). Os resultados mostraram tempos de resposta entre 0,1 s e 2,3 s por operação, com consumo de memória estável, confirmando que o sistema permanece responsável em cenários habituais de utilização pessoal.

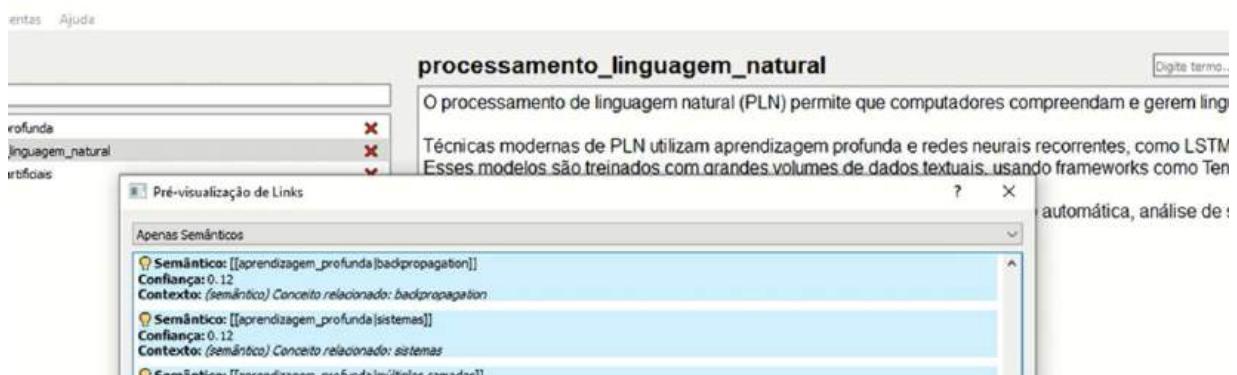


Figura 4. Pré-visualização de ligações semânticas sugeridas pelo sistema.



Figura 5. Ligações semânticas aplicadas entre notas.

5.2. Discussão

Os resultados da avaliação indicaram a viabilidade técnica da abordagem proposta pelo SmartNote e a sua capacidade de abordar as limitações identificadas nas ferramentas PKM existentes [Alles, Giozza & de Oliveira Alburquerque, 2018], [Afzal et al., 2025], [Ghadekar et al., 2023], [Gargari & Habibi, 2025]:

- *Descoberta de Relações Implícitas:* O SmartNote demonstrou de forma consistente a capacidade de identificar e criar ligações entre notas com base na sua similaridade semântica, ultrapassando a dependência de correspondências literais [Afzal et al., 2025], [Ghadekar et al., 2023]. Esta funcionalidade, evidenciada pelos testes de integração com centenas de notas, revelou conexões implícitas que seriam difíceis de descobrir manualmente, contribuindo para enriquecer a rede de conhecimento do utilizador e mitigando a fragmentação do conhecimento.
- *Escalabilidade Cognitiva:* A integração do assistente de IA baseado em RAG mostrou-se eficaz na redução da carga cognitiva [Gargari & Habibi, 2025]. Os testes de interação com o assistente indicaram que os utilizadores conseguem obter respostas contextuais e summarizadas a partir do seu corpo de notas, facilitando a compreensão e a síntese de informações complexas. O desempenho do sistema em volumes moderados de notas, com tempos de resposta inferiores a 2 segundos, confirmou a sua adequação para utilização pessoal.
- *Soberania e Disponibilidade dos Dados:* A arquitetura totalmente local do SmartNote foi validada, sugerindo que o sistema funciona sem dependência de serviços em *cloud*, reforçando a privacidade e aumentando o controlo do utilizador sobre os seus dados [Gao & Brink, 2019], [Baraku et al., 2025]. A capacidade de funcionar em modo *offline* contribui para a disponibilidade contínua do conhecimento, independentemente da conectividade à Internet.

5.3. Limitações Identificadas e Lições Aprendidas

Embora o SmartNote se tenha revelado uma prova de conceito promissora, o processo de desenvolvimento e avaliação revelou algumas limitações importantes, fornecendo lições relevantes e indicando caminhos futuros de investigação:

- *Falsos Positivos em Ligações Semânticas:* O sistema apresentou ocasionalmente sugestões de ligações de baixa relevância. Apesar de mitigadas por limiares de similaridade, estas situações evidenciam a necessidade de algoritmos mais sofisticados para desambiguação semântica e contextualização, de modo a refinar a precisão das conexões.
- *Consumo de Recursos em Grandes Volumes:* A indexação de coleções muito extensas de notas no FAISS pode exigir um consumo considerável de memória [Ghadekar et al., 2023]. Embora adequada para a maioria dos utilizadores pessoais, a escalabilidade para conjuntos de dados muito grandes exigirá otimizações adicionais, como índices FAISS aproximados ou estratégias de indexação incremental. Esta experiência sublinha a importância de equilibrar precisão com eficiência de recursos.
- *Desempenho Linguístico em Português:* Os modelos multilíngues utilizados apresentaram uma precisão ligeiramente inferior em português quando comparados com o inglês [Simplício, Semedo & Magalhaes, 2024].

- O fine-tuning de modelos específicos para a língua portuguesa ou a utilização de modelos pré-treinados em grandes corpora em português poderia contribuir para melhorar a qualidade dos embeddings, a precisão das ligações semânticas e as respostas do assistente IA local.

Estas limitações e as lições delas derivadas são relevantes para o aperfeiçoamento contínuo do SmartNote e para futuras investigações no domínio da gestão de conhecimento pessoal, permitindo uma evolução mais informada, eficaz e adaptada às necessidades dos utilizadores.

6. Conclusão

O SmartNote, disponível num repositório Git público⁶, representa um contributo para a gestão do conhecimento pessoal, abordando diretamente as lacunas deixadas pelas ferramentas PKM convencionais [Fathizargaran & Cranefield, 2017, Wright, 2005, Wang, Gao & Wang, 2010]. A sua principal contribuição reside na demonstração de um conceito exploratório: um sistema de gestão de conhecimento pessoal que é simultaneamente inteligente, escalável e privado. Ao integrar a descoberta automática de relações semânticas [Afzal et al., 2025, Ghadekar et al., 2023] e um assistente de IA local baseado em RAG [Gargari & Habibi, 2025], o SmartNote introduz uma dimensão adicional à gestão de notas, passando de uma tarefa sobretudo organizativa para um processo ativo de descoberta e interação com o conhecimento.

Em contraste com ferramentas existentes, que exigem esforço manual para criar ligações e que frequentemente comprometem a privacidade dos dados devido à dependência da *cloud* [Gao & Brink, 2019, Baraku et al., 2025], o SmartNote oferece uma alternativa consistente. A sua arquitetura totalmente local garante que o utilizador mantém o controlo total sobre os seus dados, um aspeto cada vez mais valorizado num contexto digital onde a privacidade é uma preocupação crescente. A capacidade de operar offline e a ausência de dependências externas tornam o sistema resiliente e fiável.

O sucesso na implementação das funcionalidades de ligações semânticas e do assistente de IA RAG valida a premissa de que é possível construir sistemas de conhecimento pessoal que vão além da mera organização, permitindo uma exploração mais profunda e uma interação mais intuitiva com a informação [Afzal et al., 2025, Ghadekar et al., 2023, Gargari & Habibi, 2025]. As lições aprendidas com as limitações identificadas durante o desenvolvimento, como a necessidade de desambiguação semântica e a otimização de recursos [Ghadekar et al., 2023, Simplício, Semedo & Magalhaes, 2024], fornecem um roteiro para futuras melhorias e para a evolução do SmartNote para um sistema mais sofisticado e eficiente.

O SmartNote não é apenas uma ferramenta, mas uma prova de conceito que contribui para repensar abordagens para a gestão de conhecimento pessoal, posicionando a tecnologia como um facilitador da compreensão e da síntese de ideias, em vez de um mero repositório de informações. A abordagem modular e extensível abre caminho para futuras integrações e para a adaptação a novas necessidades e tecnologias, garantindo a sua relevância contínua no campo da gestão de conhecimento.

⁶ <https://github.com/Daniseph/SmartNote>

7. Trabalho Futuro

Este trabalho apresentou o SmartNote, um protótipo exploratório que indica a viabilidade de um sistema de gestão de conhecimento pessoal capaz de mitigar algumas das limitações das ferramentas existentes. Através da sua conceção e implementação, demonstrou-se com sucesso:

- A geração automática de ligações semânticas entre notas, permitindo a descoberta de relações implícitas e contribuindo para ampliar a rede de conhecimento do utilizador.
- A interação inteligente via IA local, com um assistente baseado em RAG que fornece respostas contextuais que podem ajudar a reduzir a carga cognitiva.
- A garantia de soberania e privacidade dos dados através de um processamento 100% offline e local, conferindo total controlo ao utilizador.

O SmartNote sugere a premissa de que é possível construir sistemas de conhecimento pessoal que são simultaneamente mais inteligentes, escaláveis e privados [Afzal et al., 2025, Ghadekar et al., 2023, Gargari & Habibi, 2025]. As lições aprendidas durante o desenvolvimento, especialmente no que diz respeito à otimização de recursos e à especificidade linguística, são relevantes para a evolução futura do sistema.

Como trabalhos futuros, destacam-se as seguintes direções possíveis:

- *Fine-tuning de modelos específicos para português:* Aprimorar a precisão dos embeddings e das respostas do assistente de IA através do treino de modelos de linguagem em grandes corpora de texto em português, ou a adaptação de modelos existentes para o idioma.
- *Indexação incremental para maior escalabilidade:* Implementar mecanismos de indexação incremental no FAISS para otimizar o consumo de memória e o tempo de processamento em coleções de notas muito extensas, permitindo lidar de forma eficiente com volumes de dados maiores.
- *Visualização avançada de grafos de conhecimento:* Explorar a criação de interfaces visuais interativas que representem a rede de conhecimento do utilizador como um grafo, permitindo uma exploração intuitiva das relações semânticas e a identificação de clusters temáticos.
- *Implementação de desambiguação semântica:* Criar algoritmos que resolvam ambiguidades em termos e conceitos, reduzindo falsos positivos na geração de ligações semânticas e aumentando a precisão das conexões.
- *Integração com ferramentas PKM existentes via plugins:* Desenvolver um sistema de plugins que permita ao SmartNote integrar-se com outras ferramentas PKM populares, facilitando a importação/exportação de dados e promovendo a interoperabilidade, expandindo o seu ecossistema.

Este projeto marca o início de uma nova abordagem à organização do conhecimento pessoal, onde a tecnologia reforça as ligações entre ideias e aprofunda a forma como pensamos e aprendemos, com um foco na privacidade e no controlo do utilizador.

REFERÊNCIAS

- Clemente, B. E., & Pollara, V. J. (2005). Mapping the course, marking the trail. *IT Professional*, 7(6), 10–15. <https://doi.org/10.1109/MITP.2005.149>
- Fathizargaran, R., & Cranefield, J. (2017). Web 2.0 and personal knowledge management: A framework of skills for effectiveness. In R. Helms, J. Cranefield, & J. Van Reijse (Eds.), *Social Knowledge Management in Action* (Vol. 3, pp. 101–122). Springer International Publishing. https://doi.org/10.1007/978-3-319-45133-6_6
- Da Silva, G. S., & Ulbricht, V. R. (2024). Second brain and self-directed learning in personal knowledge management. In Á. Rocha, C. Montenegro, E. T. Pereira, J. A. M. Victor, & W. Ibarra (Eds.), *Management, tourism and smart technologies* (Vol. 1190, pp. 431–441). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-74825-7_38
- Thaul, W., Bleimann, U., & Clarke, N. (2012). Personal knowledge management beyond versioning. *Proceedings of the 12th International Conference on Knowledge Management and Knowledge Technologies*, 1–4. <https://doi.org/10.1145/2362456.2362492>
- Gao, L., & Brink, A. G. (2019). A content analysis of the privacy policies of cloud computing services. *Journal of Information Systems*, 33(3), 93–115. <https://doi.org/10.2308/isys-52188>
- Eppler, M. J., & Mengis, J. (2004). The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines. *The Information Society*, 20(5), 325–344. <https://doi.org/10.1080/01972240490507974>
- Whelan, E., & Teigland, R. (2010). Managing information overload: Examining the role of the human filter. *SSRN*. <https://doi.org/10.2139/ssrn.1718455>
- Alles, V. J., Giozza, W. F., & de Oliveira Alburquerque, R. (2018). Natural language processing to classify named entities of the Brazilian Union Official Diary. In *2018 13th Iberian Conference on Information Systems and Technologies (CISTI)* (pp. 1–6). IEEE. <https://doi.org/10.23919/CISTI.2018.8399215>
- Wright, K. (2005). Personal knowledge management: Supporting individual knowledge worker performance. *Knowledge Management Research & Practice*, 3(3), 156–165. <https://doi.org/10.1057/palgrave.kmrp.8500061>
- Wang, S., Gao, G., & Wang, X. (2010). Creating a personal knowledge management system in a learning society. In *2010 International Conference on Networking and Digital Society* (pp. 494–497). IEEE. <https://doi.org/10.1109/ICNDS.2010.5479248>
- Ackerman, M. S., & McDonald, D. W. (2000). Answer Garden 2: Merging organizational memory with collaborative help. *ACM Transactions on Information Systems*, 18(1), 32–58. <https://dl.acm.org/doi/10.1145/240080.240203>

- Pu, Y., Beck, D., & Verspoor, K. (2025). Enriched knowledge representation in biological fields: A case study of literature-based discovery in Alzheimer's disease. *Journal of Biomedical Semantics*, 16(1), 3. <https://doi.org/10.1186/s13326-025-00328-3>
- Xu, B., Wang, J., Zhao, Z., Lin, H., & Xia, F. (2024). HIM: Discovering implicit relationships in heterogeneous social networks. In ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 5875–5879). IEEE. <https://doi.org/10.1109/ICASSP48485.2024.10447273>
- Karr-Wisniewski, P., & Lu, Y. (2010). When more is too much: Operationalizing technology overload and exploring its impact on knowledge worker productivity. *Computers in Human Behavior*, 26(5), 1061–1072. <https://doi.org/10.1016/j.chb.2010.03.008>
- Edress, Z., & Ortakci, Y. (2024). Optimizing text summarization with sentence clustering and natural language processing. *International Journal of Advanced Computer Science and Applications*, 15(10), 1005-1011. <https://doi.org/10.14569/IJACSA.2024.01510115>
- Aal, K., & Rüller, S. (2025). From personal knowledge management to the second brain to the personal AI companion. In Proceedings of the 2025 ACM International Conference on Supporting Group Work (pp. 90–93). ACM. <https://doi.org/10.1145/3688828.3699647>
- Baraku, V., Ramadani, E., Paraskakis, I., Veloudis, S., & Yadav, P. (2025). Defining personal data sovereignty: An ontologically-based framework facilitating subject privacy control. *Data and Information Management*, 100108. <https://doi.org/10.1016/j.dim.2025.100108>
- Afzal, T., Abdul Rauf, S., Malik, M. G. A., & Imran, M. (2025). Fine-Tuning QurSim on Monolingual and Multilingual Models for Semantic Search. *Information*, 16(2), 84. <https://doi.org/10.3390/info16020084>
- Ghadekar, P. P., Mohite, S., More, O., Patil, P., Sayantika, & Mangrule, S. (2023). Sentence meaning similarity detector using FAISS. In 2023 7th International Conference on Computing, Communication, Control and Automation (ICCUBEIA) (pp. 1–6). IEEE. <https://doi.org/10.1109/ICCUBEIA58933.2023.10392009>
- Simplício, A., Semedo, D., & Magalhaes, J. (2024). V-Glória: Customizing large vision and language models to European Portuguese. In Proceedings of the 1st Workshop on Customizable NLP: Progress and Challenges in Customizing NLP for a Domain, Application, Group, or Individual (CustomNLP4U) (pp. 317–326). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2024.customnlp4u-1.24>
- Gargari, O. K., & Habibi, G. (2025). Enhancing medical AI with retrieval-augmented generation: A mini narrative review. *Digital Health*, 11, 20552076251337177. <https://doi.org/10.1177/20552076251337177>



Daniel José Correia Gonçalves é Licenciado em Engenharia Informática (2025). Atualmente exerce funções na área de Gestão de Acessos no Bankinter, com responsabilidades relacionadas com *Identity & Access Management* (IAM) e segurança de sistemas de informação. Possui experiência em suporte técnico e operações de TI, tendo integrado a equipa de *helpdesk* do Banco Popular, prestando apoio aos utilizadores durante o processo de transição para o Santander, e prestado suporte técnico no Banco CTT. Frequentou formação em Microsoft Systems and Network Engineering, com foco em administração de sistemas e redes, e participou em diversos projetos académicos no âmbito da engenharia informática. Os seus principais interesses incluem IAM, segurança da informação, administração de sistemas e desenvolvimento de soluções para eficiência e automação.



Leonel Caseiro Morgado, Professor Catedrático da Universidade Aberta (UAb). Leciona métodos de investigação, ambientes imersivos e programação. É Investigador Coordenador no INESC TEC, coordenando o Núcleo deste laboratório associado na UAb. É Adido para Normas da Comissão Técnica para Ambientes de Aprendizagem Imersivos da IEEE Education Society. Membro da Comissão Diretiva e do Círculo de Académicos da Immersive Learning Research Network. Faz investigação científica e desenvolvimento tecnológico sobre ambientes imersivos desde 2000, no apoio ao ensino e à aprendizagem. É autor de mais de 200 publicações científicas. Antes da carreira académica, foi diretor técnico-comercial de uma empresa de importação, distribuição e venda de hardware, terminologista da localização do Microsoft Office 97 e do Oracle InterOffice, consultor linguístico da IBM/Lotus, e para a Universidade de Trás-os-Montes e Alto Douro coordenador de equipas de desenvolvimento *Web* e de implantação de software, e gestor de equipas de combate à infoexclusão.