I. INTRODUCTION

In the 21st century, the use of electronic documents has increased with the increasing number of social media and the growth of electronic devices [1]. People now prefer electronic media because it is considered more effective and cheaper to organize data, especially since people have started to abandon the so-called paper-based process [2]. Changes to the electronic document management system have become a trend in many sectors, especially due to the need to be more efficient, cost-effective, and more environmentally friendly. Therefore, there is an increasing dependent on digital systema to replace paper documents [3]. The growth of electronic documents has a profound impact on how documents are stored, searched, and organized, both in personal and professional life [4].

On the other hand, the world of semantic search is also growing rapidly in line with the progress of Natural Language Processing (NLP) and Machine Learning (ML), which makes the document search system more relevant and accurate. Several recent studies show that this semantic model is very effective to improve search performance. For example, Tang et al. [5] Talk about how semantic embedding and ontology-based models can improve the information retrieval system, especially if ontology integration is specific to a particular field, which makes the search results more relevant and accurate. Besides, there's research from He et al. [6] which explains better semantic encoding methods for searching complex data, pointing the trend towards more sophisticated approaches in semantic search technology.

DBMS or Database Management Systems that store metadata separately from their documents have a very important role in increasing the efficiency of the search process. Well, VDBMS, or Vector Database Management Systems, has a big advantage for managing unstructured and high-dimensional data, especially for modern AI and machine learning cases. This is very different from traditional relational databases that still rely on keyword-based searches [7]. Although relational databases and documents are still widely used, they are not optimized to capture inter-sentence relationships, so sometimes the search results are inaccurate [8]. By contrast, vector databases are specifically designed to manage complex, contextually rich data, improving search relevance and efficiency [9].

This paper aims to develop a semantic search system for personal documents using vector databases, addressing the limitations of traditional search methods. The goal is to allow users to retrieve documents more efficiently and effectively by enabling semantic matching, without relying on memory of specific file names or details.