|  |  |  |  |
| --- | --- | --- | --- |
| 2024년 서울시립대학교 컴퓨터과학부 셈틀제 최종 보고서 | | | |
| **프로젝트명** | **MAFM**  **(Multi-Agent**  **File Management)** | | |
| **과제팀 이름** | cluster23 | | |
| **개발기간** | 2024. 09. 20. ~ 2024. 11. 15. | | |
| **분야** | 자유 | | |
| **셈틀제 팀 구성원** | 이름 | 박세환 | 주재원 |
| 학번 | 2020920023 | 2020920061 |
| 연락처 | 010-3171-1729 | 010-5319-9671 |

**1. 프로젝트**

* **MAFM(Multi-Agent File Management)**은 기존 파일 관리 시스템의 한계를 극복하고 보다 **효율적인 파일 검색 기능**을 제공하는 지능형 파일 관리 시스템이다. 기존 시스템은 키워드 기반 검색에 의존하여, 사용자가 입력한 정확한 워딩이 포함된 파일만을 찾을 수 있어 검색 시간이 오래 걸리고 검색 정확도가 떨어진다. 예를 들어, 사용자가 "이전에 다운로드한 AI 관련 논문"을 찾으려고 할 때 사용자는 "AI"와 "논문" 이라는 키워드로 시스템에서 파일을 찾는다. 기존 시스템은 파일명이나 내용에 "AI"와 "논문"이라는 단어가 정확히 포함된 파일만을 검색 대상으로 삼기 때문에 논문의 제목이나 내용이 다르게 표현된 경우에는 검색 결과에서 제외될 수 있다. 만약 파일의 제목과 내용에서 "AI"와 "논문" 이라는 단어를 사용하지 않고 "인공지능"과 "연구" 라는 단어를 사용했다면 이 파일은 검색 결과에서 제외된다. MAFM은 이러한 문제를 해결하기 위해 **Multi-Agent** 구조와 **VectorDB**를 도입한다.
* 각 Agent는 하나의 디렉토리와 해당 디렉토리에 대한 VectorDB를 관리하며, 사용자의 검색 쿼리에 따라 연관성이 높은 파일을 빠르고 정확하게 찾는다. VectorDB는 단순 키워드 일치가 아닌, 벡터화된 쿼리와 파일의 거리 기반으로 검색 결과를 도출한다. 이를 통해 AI와 VectorDB에 관한 파일을 찾을 때, 해당 주제에 대한 서술이 포함된 파일을 손쉽게 찾을 수 있다. MAFM 시스템에 "이전에 다운로드한 AI 관련 논문"을 쿼리로 질의하면, MAFM은 단순히 쿼리와 정확하게 일치하는 파일을 찾는 것이 아니라 해당 쿼리 자체와 연관성이 높은 파일을 검색한다. 또한, 여러 Agent가 병렬적으로 검색을 수행함으로써 검색 성능이 크게 향상된다.

**2. 배경**

**가. 개발 배경**

* 현재 많은 파일 관리 시스템은 사용자가 파일을 찾기 위해 입력한 정확한 키워드를 기반으로 검색을 수행한다. 하지만, 사용자가 입력한 키워드와 파일에 저장된 정보가 완전히 일치하지 않는 경우, 원하는 파일을 찾기 어려운 문제가 발생한다.
* MAFM은 이러한 문제를 해결하기 위해 개발되었다. 이 시스템은 Multi-Agent 구조와 VectorDB를 도입하여, 단순 키워드 일치가 아닌 파일의 의미적 연관성을 분석하여 검색을 수행한다. 특히 MAFM은 단순히 각 디렉토리를 관리하는 에이전트들만 있는 것이 아니라, 사용자의 쿼리에 따라 적절한 디렉토리를 선택하는 에이전트와 검색 결과를 종합하는 에이전트도 존재한다. 예를 들어 사용자가 "이전에 다운로드한 AI 관련 논문"을 찾는다면, 디렉토리를 선택하는 에이전트에게 사용자의 질의가 전달이 된다. 디렉토리 선택 에이전트는 사용자가 다운로드한 파일이 위치할 가능성이 큰 "Download" 디렉토리를 관리하는 에이전트에게 질의를 전달해 “Download”폴더의 VectorDB에 검색을 진행한다. 이러한 방식으로 다중 에이전트들이 협력하여 더 효율적인 파일 검색을 수행한다.

**나. 개발 효과**

* MAFM을 도입함으로써 **파일 검색의 정확도와 속도**가 크게 향상될 것이다. MAFM은 각 에이전트가 담당하는 디렉토리에 대해 VectorDB를 유지하며, 사용자의 쿼리를 벡터로 변환하여 관련성이 높은 파일을 탐색한다. 이 과정에서 각 파일은 미리 임베딩되어 VectorDB에 저장되며, 검색은 쿼리 벡터와 파일 벡터 간의 거리 계산을 통해 이루어진다. 이를 통해 키워드 일치 여부에 상관없이 주제나 내용과 연관된 파일을 검색할 수 있어 성능이 증진된다.
* MAFM은 **키워드에 종속되지 않는 파일 검색** 기능을 제공한다는 점에서 기존 시스템과 차별화된다. 사용자는 특정 키워드 대신 주제나 내용을 기반으로 파일을 검색할 수 있어, 보다 유연한 검색이 가능하다. 예를 들어, 영어로 된 파일을 찾고자 할 때도 한국어로 질의를 입력하면 연관된 파일을 찾아낼 수 있다. 이는 MAFM의 벡터 기반 검색 방식 덕분에 가능하며, 다국어 검색을 지원함으로써 다양한 언어로 작성된 파일들도 효과적으로 검색할 수 있게 된다.
* 다양한 주제와 **많은 파일을 관리**해야 하는 환경에서 MAFM은 특히 유용하다. 연구자나 학생들이 정확한 파일명을 기억하지 못하더라도, VectorDB를 기반으로 연관된 파일을 쉽게 찾을 수 있으며, 디렉토리 선택 에이전트가 적절한 디렉토리를 우선적으로 검색하여 시간 낭비를 최소화한다. 이를 통해 파일 관리의 편의성이 높아지고, 연구나 프로젝트 작업의 효율성 또한 증가할 것이다.

**3. 설계**

텍스트, 스크린샷, 도표, 평면도이(가) 표시된 사진

자동 생성된 설명

* **Multi-Agent 구조**  
   위 다이어그램에서 나타낸 바와 같이, MAFM은 여러 에이전트가 협력하여 파일 검색과 관리를 수행한다. 각 디렉토리에는 개별적으로 담당하는 에이전트가 있으며, 사용자의 검색 요청에 따라 적절한 디렉토리를 선택하고 해당 디렉토리의 VectorDB에서 검색을 수행하는 역할을 맡고 있다. 디렉토리 선택을 조율하는 에이전트는 사용자의 쿼리와 연관된 디렉토리를 선택하여 검색 작업을 할당하며, 검색이 완료되면 결과 종합 에이전트를 호출하여 결과를 모아 사용자에게 반환한다. 이러한 구조는 병렬적으로 검색 작업을 처리할 수 있어 성능을 크게 향상시킨다.
* **핵심 알고리즘:** 
  + **파일 저장 알고리즘**
    - **파일 생성 및 저장**: 사용자가 새로운 파일을 생성하고 저장하면, MAFM은 그 파일을 처리하기 위한 작업을 시작한다.
    - **파일 분할**: 파일이 너무 클 경우, Embedding 작업을 원활하게 수행하기 위해 파일을 2048 Bytes 단위로 분할한다.
    - **파일 내용 읽기**: 분할된 파일의 내용을 차례대로 읽어 들인다.
    - **Embedding 작업**: 파일 내용을 Embedding하여 벡터로 변환한다. 이때, 하나의 파일에서 분할된 여러 파일들은 같은 ID를 공유해야 한다.
    - **VectorDB 저장**: 변환된 벡터를 해당 디렉토리의 VectorDB에 저장한다.
    - **변경사항 저장**: 디렉토리에 변경 사항이 발생했으므로 이를 기록하여 업데이트한다.
  + **파일 검색 프로세스**
    - **쿼리 전달**: 사용자가 파일을 검색할 때, 검색 쿼리를 디렉토리 선택 에이전트에게 전달한다.
    - **검색 수행**: 각 에이전트는 자신의 담당 디렉토리의 VectorDB에서 쿼리와 연관된 파일을 검색한다.
    - **검색 결과 반환**: 검색된 파일들의 경로(file path)를 반환한다.
    - **Symbolic Link 생성**: 반환된 파일 경로를 기반으로 심볼릭 링크(Symbolic Link)를 생성한다.
    - **임시 디렉토리 저장**: 검색된 파일이 여러 개일 수 있으므로 생성된 심볼릭 링크들을 임시 디렉토리(temp)에 모아 저장한다.
    - **사용자에게 표시**: 최종적으로 임시 디렉토리를 사용자에게 보여준다.

이러한 검색 및 저장 과정은 MAFM의 Multi-Agent 구조에서 병렬적으로 처리되어, 다수의 파일을 빠르게 검색하고 결과를 제공하는 데 최적화되어 있다.

**4. 기능**

* **기능**
  + **핵심적인 기능을 소개.**
    - 프로그램의 코드를 설명.
    - 구현 알고리즘, 실행 방법 등을 구체적으로 설명.
  + **1, 2학년 학생들도 이해할 수 있도록 쉽게 설명한다.**

**5. 결과 및 평가**

* **결과 및 평가**
  + **개발한 기술 또는 작품을 평가하는데 필요한 평가항목 및 평가방법, 개발목표치 등을 기재.**