



2025학년도 1학기 SW 캡스톤디자인 경진대회

인공지능 탐색기를 활용한 저장소 최적화 기술 개발

팀 명 CV마스터
지도교수 김윤경

팀 원 김은혜(컴퓨터인공지능학과, 202212112), 박태호(컴퓨터인공지능학과, 202022164), 최준호(컴퓨터인공지능학과, 202219456), 이다은(컴퓨터인공지능학과, 202011696)
산업체 (주)올포랜드

개발 동기 및 목적

현업 환경에서는 문서 파일이 반복적으로 생성되거나 복사되면서, 유사한 이름과 내용의 파일이 동일 폴더에 다수 저장되는 경우가 흔하게 발생합니다. 특히 업무에서 사용하는 양식, 보고서, 명세서 등은 이름만으로는 구분이 어렵고, 수정일자나 크기만으로도 최신 파일을 식별하기 어렵습니다. 이러한 파일이 쌓이게 되면 저장소의 공간을 비효율적으로 차지하게 되며, 실제 필요한 파일을 찾는 데에도 많은 시간이 소요됩니다. 이러한 문제를 해결하기 위해 본 프로젝트는 사용자의 유휴 시간 동안 자동으로 중복 및 구버전 파일을 탐지하고 정리할 수 있는 시스템을 개발하는 것을 목적으로 합니다. 사용자의 활동 패턴을 분석하여 유휴 상태를 판단하고, 이 시점을 활용해 파일 정리를 수행함으로써 사용자의 개입 없이도 저장소를 효율적으로 관리할 수 있도록 설계되었습니다. 또한 AI 기술을 활용하여 정리뿐만 아니라 문서 분류까지 자동화함으로써, 파일 관리의 정확성과 편의성을 동시에 향상시키는 것을 목표로 하였습니다.

KLIP-SYS-AN-050_(CMM)유스케이스명세서.hwp	E#2...	86 KB	2021-01-11 오후 2:06
KLIP-SYS-AN-050_(CMM)유스케이스명세서.hwp	E#2...	86 KB	2021-01-11 오후 2:06
KLIP-SYS-AN-050_(CMM)유스케이스명세서.hwp	E#2...	86 KB	2021-01-11 오후 2:06
KLIP-SYS-AN-050_(CMM)유스케이스명세서.hwp	E#2...	85 KB	2020-10-05 오후 5:17
KLIP-SYS-AN-050_(CMM)유스케이스명세서.hwp	E#2...	183 KB	2021-01-18 오전 10:16
KLIP-SYS-AN-050_(CMM)유스케이스명세서.hwp	E#2...	86 KB	2021-01-18 오전 10:16
KLIP-SYS-AN-050_(CMM)유스케이스명세서.hwp	E#2...	86 KB	2021-05-12 오후 3:18
KLIP-SYS-AN-050_(CMM)유스케이스명세서.hwp	E#2...	86 KB	2021-01-11 오후 2:06
KLIP-SYS-AN-050_(CMM)유스케이스명세서.hwp	E#2...	85 KB	2021-06-24 오전 10:00
KLIP-SYS-AN-050_(CON)유스케이스명세서.hwp	E#2...	118 KB	2021-01-11 오후 11:01
KLIP-SYS-AN-050_(CON)유스케이스명세서.hwp	E#2...	118 KB	2021-01-11 오후 11:01
KLIP-SYS-AN-050_(CON)유스케이스명세서.hwp	E#2...	118 KB	2021-01-11 오후 11:01
KLIP-SYS-AN-050_(CON)유스케이스명세서.hwp	E#2...	84 KB	2020-10-05 오후 5:17
KLIP-SYS-AN-050_(CON)유스케이스명세서.hwp	E#2...	147 KB	2021-01-18 오전 10:16
KLIP-SYS-AN-050_(CON)유스케이스명세서.hwp	E#2...	118 KB	2021-01-18 오전 10:16
KLIP-SYS-AN-050_(CON)유스케이스명세서.hwp	E#2...	118 KB	2021-05-12 오후 3:18
KLIP-SYS-AN-050_(CON)유스케이스명세서.hwp	E#2...	118 KB	2021-01-11 오후 11:01

주요 기술

본 프로젝트에서는 LSTM(Long Short-Term Memory) 기반 시계열 모델을 사용하여 사용자 활동 데이터를 분석하고 유휴 상태를 예측하였습니다. 해시값(SHA)을 활용한 중복 파일 탐지, 파일명 및 내용 기반의 유사도 분석을 통해 구버전 파일 정리 기술도 적용하였습니다. 또한 LLM 기반 자연어 처리 기술을 활용하여 파일명을 분석하고 본문 키워드까지 추출하여 주제 분류를 수행하였습니다. 예측 모델은 사용자별 유휴 기준값에 맞춰 작동하며, 실시간 입력 데이터를 기반으로 향후 유휴 여부를 5분 단위로 세분화하여 예측합니다. 정리 기술은 해시 비교와 텍스트 유사도 연산을 병행하여 높은 정확도로 중복·구버전 파일을 선별할 수 있도록 구현하였습니다. 분류 기술은 문서 명칭이 불분명한 경우에도 내용 분석을 통해 핵심 키워드를 추출하고 카테고리를 결정합니다. 이러한 기술들을 통합하여 사용자 개입 없이도 예측 → 정리 → 분류까지 자동으로 연계 작동하는 시스템을 완성하였습니다. 또한 LSTM과 LLM을 각 기능에 특화되도록 분리 적용함으로써 시스템 전체의 처리 속도와 안정성을 확보하였습니다. 각 기술 요소는 Python, TensorFlow, Pandas 등 오픈소스 기반 기술을 활용하여 유연성과 확장 가능성을 고려하여 개발하였습니다. 전반적으로 기술 구성은 실시간성과 자동화에 중점을 두었으며, 실제 유휴 시간 활용을 극대화하는 방향으로 설계되었습니다.

개발 내용

본 시스템은 사용자의 실제 컴퓨터 사용 패턴을 반영하여 자동으로 작동되는 구조로 설계되었습니다. 먼저 사용자 환경에 따라 자원 사용량과 입력 활동을 기반으로 유휴 상태를 식별할 수 있도록, 관련 데이터를 일정 시간 동안 수집하는 기능을 구현하였습니다. 이후 실시간 입력값을 기반으로 시계열 LSTM 모델을 이용하여 향후 일정 시간 동안의 유휴 여부를 예측하고, 해당 시간 동안 자동 정리 기능이 작동될 수 있도록 연계하였습니다. 정리 단계에서는 동일 해시값을 가진 중복 파일을 탐지하고, 최신 파일 1개만 남긴 뒤 나머지는 격리하였습니다. 파일명이 유사하거나 내용이 거의 같은 문서들은 구버전 판단 기준에 따라 선별적으로 제거되었습니다. 이후 LLM 기반 분류 모듈이 작동하여, 파일명 분석 → 내용 분석 순으로 카테고리를 판단하고, 연도·분기·주제별 폴더를 자동 생성하였습니다. 정리 및 분류는 유휴 시간 동안 자동으로 실행되어 사용자의 작업 흐름에 영향을 주지 않도록 처리되었습니다. 또한 분류 후 생성된 폴더 구조는 직관적으로 정리되어 문서 탐색 및 관리 효율성을 높이는 데 기여하였습니다. 기능별 처리는 단계적으로 작동되며, 예외 상황에 대한 기본적인 오류 처리 로직도 포함하여 안정성을 확보하였습니다. 전체 개발은 모듈화된 구조로 구성되어, 추후 다양한 파일 확장자 지원 및 기능 확장을 용이하게 만들었습니다.

결과 및 분석

프로그램 최초 실행 시, 사용자 개별의 유휴 상태 판단을 위한 기준 데이터를 수집하였습니다. CPU 및 디스크 사용량, 마우스·키보드 입력 횟수를 바탕으로 유휴 기준값을 자동 설정하였으며, 이후 실시간 데이터를 기반으로 최근 20분간의 활동을 분석하여 향후 30분간의 유휴 여부를 5분 간격으로 예측하도록 구성하였습니다. 예측 결과 평균값이 0.6 이상일 경우 유휴 상태로 판단하였고, 그 시점부터 자동으로 파일 정리 기능이 작동하도록 설계하였습니다. 파일 정리 과정에서는 해시값(SHA)을 이용해 동일 파일을 식별하고, 버전이 낮거나 오래된 파일을 자동으로 격리하였습니다. 이후 LLM 기반 분류 시스템을 통해 파일명 및 본문 내용을 분석하여 연도, 분기, 주제별로 분류·정리하는 방식으로 디렉토리 구조를 자동 구성하였습니다. 이를 통해 사용자의 개입 없이도 유휴 시간 동안 효과적인 파일 정리와 구조화가 가능함을 확인하였습니다. 중복 파일의 경우 동일한 해시값을 가진 파일만 선별하여 가장 최근 수정된 파일을 남기고 나머지를 삭제 처리하였습니다. 구버전 파일은 유사한 이름과 내용 기반 비교를 통해 버전이 낮은 문서를 식별하였으며, 분류 작업에서는 파일 이름이 불분명한 경우에도 내부 텍스트를 분석해 키워드를 추출하고 적절한 주제로 분류가 이루어졌습니다. 이러한 작업들은 유휴 시간 중 자동으로 실행되며, 사용자의 작업 흐름에 영향을 주지 않는 선에서 처리되도록 구현되었습니다.