



금융핀

≡ 팀 번호	4팀(C4)
≡ 팀장	구경덕
≡ 팀원	지은현 이하얀 구경덕
≡ 데이터	https://developers.naver.com/docs/serviceapi/search/news/news.md
📎 제출 자료(Git)	https://github.com/hayannn/FinPin_LangchainTon
≡ 참고	https://miro.com/app/board/uXjVlY1GRfk=/

주제

금융 뉴스 요약 프로젝트

▼ 아이디어

- 헌법을 포함한 다른 법안들까지 포함하는 챗봇
 - 법률 자문 등의 상담 비용 절감
- 금융 뉴스 요약 챗봇
 - 헤드라인 모음
- 주야간(시간대)에 따른 이용가능한 병원 정보를 알려주는 프로젝트
 - ...
- 추가 상의
 - Miro 링크
 - https://miro.com/welcomeonboard/cGUyN3B1NFBUciRvb3B4WnBZNmQ5MzEwVFdYNTdKN0EweENxSzBiRmdTb21GaTlqZGxwYjZmYm5z?share_link_id=465805835613



팀명 선정

금융핀

- 금융 + 핀(Pin: 중요한 부분을 집어낸다는 의미, finance: 금융 → 중의적 표현)

PoC

1. 문제 정의
2. 목표 & 가설
3. RAG 구조 개요
4. 데이터 & 기술 스택
5. 검증 방법
6. 구현 시 어려움 & 해결 방안
7. 향후 발전 가능성
8. 역할 분담
9. Appendix
 - 페르소나 시나리오
 - 사용자 시나리오 다이어그램
 - 서비스 아키텍처
 - WBS
 - 참고 링크

1. 문제 정의

- 주제
 - 금융 뉴스 요약 프로젝트
 - 여러 뉴스 사이트에서 제공하는 데이터셋(네이버, 구글 등)을 기반으로 금융 뉴스를 요약
- 문제 정의
 - 여러 사이트들을 보면서 자료를 수집하는 데에 시간적인 낭비가 심함
 - ChatGPT의 요약 도움
 - 제대로 잘 요약되지 않는 부분이 많아 불편함을 자주 느낌
 - 출처 등이 명확하지 않아 다소 신뢰도가 떨어짐

2. 목표 & 가설

- 목표
 - MVP이기 때문에 특정 플랫폼 1가지를 선정하여, 해당 플랫폼의 뉴스 기사 API or 데이터셋 등을 제공받아 요약
 - 사용자가 더욱 쉽게 경제 뉴스를 접할 수 있고, 빠르게 신뢰성있는 정보를 파악할 수 있도록 하는 것이 목표
- 가설
 - 뉴스 요약을 통해 빠르게 정보 파악 가능
 - 시간 및 비용 절감을 위한 사용자에게 적합할 것

3. RAG 구조 개요

- API 이용
 - 네이버 뉴스 기사 검색 API를 통한 기사 수집
 - 데이터 분할
 - 데이터 임베딩
 - 임베딩 검색
 - 벡터 스토어에 임베딩된 정보 저장
 - 사용자 프롬프트로 쿼리를 받음
 - 답변 생성
- PDF 파일 이용

- 문서 업로드
- 문서 분할
- 문서 임베딩
- 임베딩 검색
 - 벡터 스토어에 임베딩된 정보 저장
 - 사용자 프롬프트로 쿼리를 받음
- 답변 생성

4. 데이터 & 기술 스택

- 데이터 : 네이버 뉴스 기사 검색 API
 - API 사용이 어려울 경우, 링크 혹은 PDF 파일 이용하는 방법으로 교체
- 기술 스택(추후 추가 예정)
 - Naver Search API
 - OpenAI API
 - LangChain
 - PyPDFLoader
 - FAISS(Vector Store) or Chroma
 - Streamlit

5. 검증 방법

- 특정 뉴스 기사를 기준으로 기본 ChatGPT 모델 요약 답변과 프로젝트 요약 답변을 비교하는 방식으로 검증
 - 혹은 동일한 질문을 바탕으로 답변 내용을 비교
 - 유사한 질문을 계속해서 던질 경우, 동일한 답변을 내는지 여부

6. 구현 시 어려움 & 해결 방안

- API 사용
 - API 사용을 1순위로 두되, 사용이 어려울 경우 링크 첨부 or 문서 파일을 사용하는 방식으로 해결할 예정
- API 데이터 부족
 - 검색 결과가 없을 경우, 사용자에게 대체 기본 메시지를 표시하거나 결과가 없다는 내용을 표시할 예정
- VectorStore
 - FAISS 사용 시 GPU 등의 성능 이슈가 발생할 수 있어 다른 vectorstore(Chroma 등)도 사용 및 테스트를 통해 사용 가능한 vectorstore를 이용할 예정
- 사용자 입력 처리
 - 추후 논의가 더 필요한 사항이지만, 기본적인 오타 또는 비정형 입력 등에 대한 처리가 있어야 할 것으로 보임
 - 공백 제거, 오타 수정 등의 로직을 추가할 예정(시간이 되는 선에서)

7. 향후 발전 가능성

- 기존 ChatGPT vs API를 넣은 모델 vs 문서 파일을 넣은 모델 3가지로 요약 성능 비교
 - API 모델과 문서 파일 모델 중 가장 좋은 요약 답변을 내는 모델을 사용하여 확장
- API 종류 확장
- 요약 내용에 기초 금융 용어에 대한 설명을 추가해주는 것
 - 단순 웹 검색이 아닌 신뢰성 있는 정보에서 추출해주는 방식
- 금융뿐만 아니라 다른 분야에 특화된 모델로도 발전 가능

8. 역할 분담

- 데이터 및 기능 관련 레퍼런스 서치 : 공동
- 기능 구현 : 공동
- 자료 작성 및 정리(발표 자료 외) : 이하얀
- 발표 : 구경덕

9. Appendix.

페르소나 시나리오

페르소나: 일반 사용자

배경

- 이름: 박수진
- 직업: 사회초년생
- 연령대: 20대 후반
- 기술 수준: 보통 (검색엔진 등의 사용 정도에 익숙한 정도)
- 관심사: 최신 경제 뉴스, 금융 트렌드, 투자 정보 검색
- 목표: 빠르고 정확한 금융 뉴스를 제공받기를 원함

• 경제 뉴스로 업무와 개인 투자 역량을 키우고자하는 사회초년생

- **상황:** 신입사원인 박수진은 IT 스타트업에서 마케팅 업무를 담당하고 있다. 회사는 기술 제품과 서비스를 제공하지만, 최근 경제 상황에 따라 소비자 트렌드가 변화하고 있어 최신 경제 뉴스를 신속히 파악하는 것이 중요하다. 동시에, 박수진은 개인적으로 주식과 ETF에 관심이 많아 경제 뉴스를 통해 투자 기회를 파악하고 싶어 한다. 하지만 바쁜 업무 속에서 정보를 효율적으로 정리하고 분석하는 데 어려움을 겪고 있다.

• 사용 시나리오:

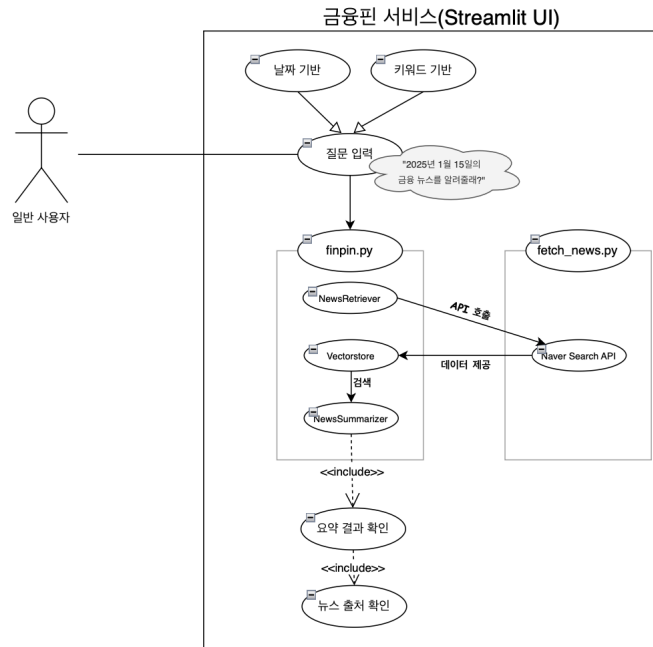
- 출퇴근 시간에 버스나 지하철을 기다리면서 새벽동안, 혹은 업무시간동안 나왔던 경제기사를 알려달라고 금융핀 서비스에 접속해서 "2025년 1월 10일의 소비자 트렌드 관련 경제 뉴스"를 검색한다.
- 서비스는 관련 기사의 핵심 내용을 요약하고, 트렌드 분석과 관련 투자 기회를 보여준다.
예를 들어, 서비스는 소비 트렌드 변화와 관련된 주요 산업과 기업 정보를 제공하며, 뉴스 내용과 연관된 투자 종목(주식, ETF 등)에 대한 추가 정보를 학습한다.
- 요약된 내용을 바탕으로 업무에서 인사이트를 얻는 동시에, 개인적으로 관심 있는 투자 영역을 분석해 자신의 투자 전략을 구체화한다.

• 목표:

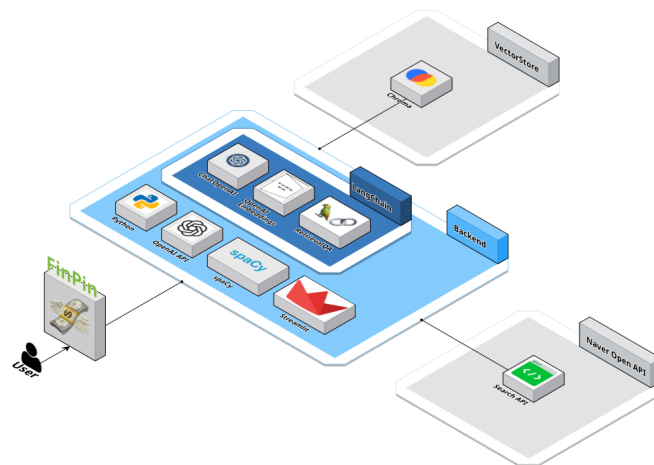
- 업무 효율성을 높이기 위해 주요 경제 뉴스와 소비자 트렌드를 신속히 파악하고 싶다.
- 뉴스 요약을 기반으로 개인 투자 아이디어를 구체화하고, 투자 분석 능력을 키우고 싶다.
- 시간을 절약하며 신뢰할 수 있는 정보를 통해 업무와 개인 목표를 동시에 달성하고 싶다.

사용자 시나리오 다이어그램

- 최종본



서비스 아키텍처



WBS







Day 01

- 날짜 : 1월 10일 (금)
- 팀장 선정 및 팀명 정하기
- 아이디어 선정
- PoC 노선 페이지 작성
- 페르소나 시나리오, 사용자 시나리오 다이어그램, 서비스 아키텍처, WBS 작성

Day 1.5

- 날짜 : 1월 11일 (토) ~ 1월 12일(일)
- 네이버 뉴스 기사 API 연동 및 간단 코딩




Day 02

- 날짜 : 1월 13일 (월)
- 팀별 주제 공유(오전)
- 랭체인톤 진행(팀별 주제 공유 이후)
 - API 연동 관련 코드 작성
 - 세부 코드 작성 및 트러블 슈팅
 - OpenAI API Key 연동 
 - OpenAI Embedding(뉴스 데이터의 벡터화) 
 - 뉴스 데이터 전처리 
 - 청크 분할(긴 기사, 기준은 ~~500자~~ 1000자)
 - 불필요한 HTML 태그 제거(API 문서 응답 예시 참고)
 - 벡터스토어 생성 & 검색 
 - 벡터화 : OpenAI 임베딩 모델을 이용해 기사 내용 벡터화
 - 벡터 스토어 : ~~FAISS~~ Chroma
 - 사용자 질문과 유사한 기사 검색
 - 체인을 이용해 요약 생성(최적 응답) 
 - 검색한 기사를 이용해 OpenAI LLM 모델을 이용해서 핵심 내용만 추출
 - 체인 종류 : RetrievalQA, ~~stuff~~ 등
 - 결과 
 - 생성된 요약 제공 + 링크 제공

▼ 중간 코드 공유

경덕님, 은현님 코드

Day 03

- 날짜 : 1월 14일 (화)
- 랭체인톤 진행(오전)
 - 코드 완성 및 최종 결과물 영상 찍기 
- 발표 자료 구성(오후) 
- 깃허브 레포지터리 구성(오후) 

PPT 내용 구성

Day 04


- 날짜 : 1월 15일 (수)
- 발표 준비
 - 오전 10:00 ~ 12:00 - 발표 연습 및 피드백
- 결과물 및 발표 자료 제출
 - ~ 12:00까지
- 발표 및 회고(오후)

참고 링크

- Naver Developers

NAVER Developers

네이버 오픈 API들을 활용해 개발자들이 다양한 애플리케이션을 개발할 수 있도록 API 가이드와 SDK를 제공합니다. 제공중인 오픈 API에는 네이버 로그인, 검색, 단축URL, 캠퍼를 비롯 기계번역, 음성인식, 음성합성 등이 있습니다.


 <https://developers.naver.com/main/>



- Naver News Search API

검색 > 뉴스 - Search API

검색 > 뉴스 뉴스 검색 개요 개요 사전 준비 사항 뉴스 검색 API 레퍼런스 뉴스 검색 결과 조회 오류 코드 검색 API 뉴스 검색 구현 예제 뉴스 검색 개요 개요 사전 준비 사항 개요 검색 API와 뉴스 검색 개요 검색 API는 네이버 검색 결과를 뉴스, 백과

 <https://developers.naver.com/docs/serviceapi/search/news/news.md>

▼ 주제 관련 기타 참고

- NewsAPI
 - <https://newsapi.org/>
- 주제 관련 참고 레퍼런스(코드 참고가 아닌 주제 참고)
 - 내부 파일
 - [AI 실시간 뉴스 요약](#)
 - 크롤링
 - [취업 준비생을 위한 직무별 뉴스 요약 서비스](#)
 - [GPT를 활용한 주식 뉴스 AI 요약 프로젝트](#)
 - API 사용
 - [뉴스 요약 챗봇 포뉴](#)
 - 전체 플로우 참고
 - [Alsenet](#)