

# 2025년 컴퓨터정보공학부 졸업작품 최종보고서

프로젝트명	NLP 기반 뉴스 경제 용어 분석 및 학습 시스템			
팀명	STJ			
팀원 현황	학번	성명	캡스톤설계 교과목 이수학기	비고
	2020202037	염정호	2025-1	
	2020202095	김준원	2025-1	
	2020202008	박세영	2025-1	
	2020202006	최현서	2025-1	

신청일자 : 2025 . 11 . 28 .

신청자(팀장) : 염정호 (인)

지도교수 확인 : 신동화 (인)



**광운대학교**  
KwangWoon University

# 목 차

1. 과제의 개요 .....	1
가. 배경 및 필요성 .....	1
나. 목표 .....	1
다. 개발 내용 .....	1
2. 과제의 내용 .....	2
가. 설계 및 개발의 내용 .....	2
나. 수행 방법 및 추진 과정 .....	2
다. 최종 결과물 .....	2
라. 개선 방안 .....	2
3. 과제의 향후 계획 .....	3
가. 활용 방안 .....	3
나. 기대 효과 .....	3
4. 참고문헌 .....	4
5. 별첨 .....	4

## 1. 과제의 개요

### 가. 배경 및 필요성

최근 금융 시장의 불확실성과 디지털 금융 환경의 확산으로 인해 일반 대중의 금융 정보에 대한 접근성과 이해도가 더 중요 해지고 있다. 그러나 국내 금융 문해력 수준은 낮은 편으로, 2018년 S&P 조사에 따르면 한국 성인의 금융 문맹률은 67%에 달하며, 2022년 금융감독원의 디지털 금융 이해력 조사에서도 평균 점수는 42.9점으로 나타났다. 이러한 낮은 금융 문해력은 경제 뉴스의 활용에도 직접적인 장애 요소가 된다. 경제 뉴스는 다양한 경제·금융 용어와 복잡한 개념을 포함하고 있으며, 일반 독자에게는 낯설고 어려운 내용이 많아 정보 격차를 심화시키고 있다. 기존 뉴스 플랫폼은 요약이나 추천 기능은 제공하지만, 전문 용어에 대한 충분한 해설이나 개인 맞춤형 학습 기능은 부족한 실정이다. 따라서, 누구나 경제 뉴스를 쉽게 이해하고, 나아가 금융 지식까지 체계적으로 습득할 수 있는 플랫폼의 필요성이 대두되고 있다.

### 나. 목표

본 과제는 금융 지식이 부족한 일반 사용자도 경제 뉴스를 쉽게 이해하고, 나아가 경제·금융 용어와 개념을 자연스럽게 학습할 수 있도록 돕는 플랫폼을 개발하는 것을 목표로 한다. 이를 위해 첫째, 기사 내 등장하는 경제·금융 용어나 경제 개념에 대해 사용자가 클릭만으로 쉽고 직관적인 해설을 볼 수 있는 '경제 뉴스 해석 지원 시스템'을 구축하고자 한다. 이는 사용자가 뉴스 콘텐츠를 소비하는 과정에서 실시간으로 용어를 이해하고 금융 지식을 쌓을 수 있도록 도와준다. 둘째, 사용자가 이해하지 못하거나 관심을 가지는 경제·금융 용어를 선택해 저장하고, 이를 기반으로 개인의 수준에 맞춘 맞춤형 학습 콘텐츠를 제공하는 기능도 함께 개발한다. 이 과정을 통해 사용자는 단편적인 정보 습득을 넘어 체계적인 금융 개념 학습이 가능해진다. 셋째, 학습에 대한 흥미를 높이기 위해 퀴즈나 레벨 업과 같은 게이미피케이션 요소를 도입하여 반복 학습을 유도하고 지속적인 플랫폼 활용을 장려한다. 또한, 인공지능 기술을 활용한 뉴스 자동 요약 기능과 유사 키워드 기반의 뉴스 추천 기능을 통해 정보 탐색 효율성을 높이며, 실제 기사에 등장하는 경제·금융 용어들을 체계적으로 수집해 경제·금융 용어 사전을 구축함으로써 신뢰성 있는 학습 기반도 함께 마련할 계획이다.

## 다. 개발의 내용

시스템은 사용자가 경제 뉴스를 더욱 쉽게 이해하고 금융 지식을 효과적으로 습득할 수 있도록 설계된 통합 플랫폼이다. 사용자가 뉴스 콘텐츠를 클릭하면, 서버는 해당 뉴스 본문에서 경제·금융 관련 용어를 자동으로 추출하고, 이를 기반으로 경제·금융 용어 데이터베이스와 연동하여 해당 단어의 정의와 예제를 가져온다. 이때 뉴스 본문 내 추출된 용어들은 시각적으로 구분되도록 표시되며, 사용자는 단어를 클릭하여 팝업 형식으로 해설을 확인할 수

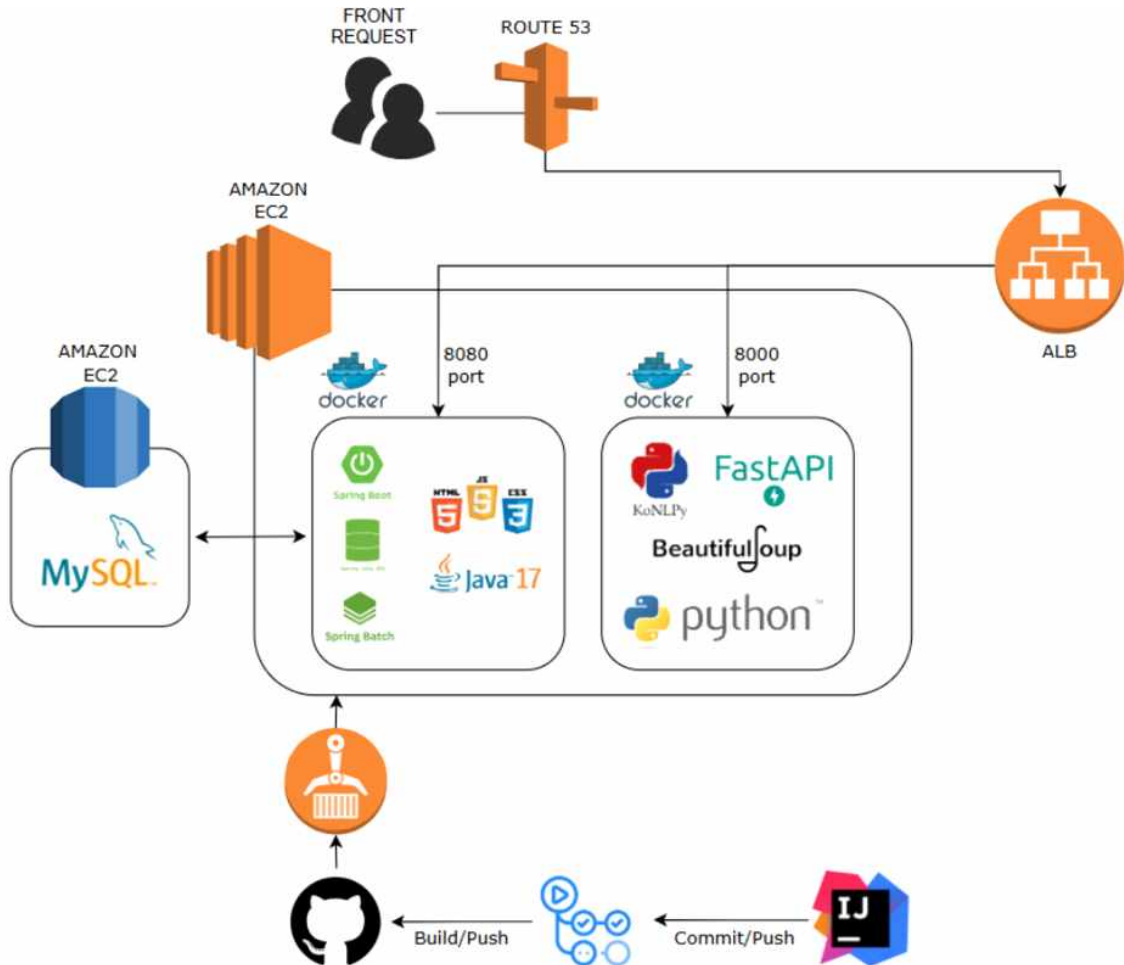
있다. 이와 함께, 사용자가 선택하거나 체크한 단어는 개인의 학습 기록에 저장되고, 시스템은 이를 분석하여 사용자 수준에 맞춘 학습 콘텐츠를 자동으로 구성한다. 학습 콘텐츠는 난이도 별로 구성되며, 퀴즈와 같은 학습 도구를 통해 게이미피케이션을 적용함으로써 사용자의 학습 몰입도를 높이는 데 중점을 둔다.

또한, AI를 활용한 뉴스 요약 기능은 사용자가 긴 경제 기사를 간략하게 파악할 수 있도록 지원하고, 뉴스 내 주요 키워드를 중심으로 유사한 뉴스 콘텐츠를 추천하여 정보 탐색의 연속성과 심화 학습을 가능하게 한다. 마지막으로, 사용자가 검색하거나 클릭한 단어 중 데이터베이스에 등록되지 않은 경우에는 생성형 AI를 통해 실시간으로 해설을 생성하며, 경제·금융 용어일 경우에는 해당 정보를 데이터베이스에 자동으로 업데이트해 지속적으로 용어 사전의 완성도를 높여 나간다. 이러한 전반적인 시스템 흐름을 통해 사용자는 단순한 기사 소비를 넘어, 실질적인 금융 학습과 정보 활용 능력을 동시에 갖출 수 있게 된다.

## 2. 과제의 내용

### 가. 설계 및 개발의 내용

#### 1) 개념 설계 (구조 설계)



본 시스템은 AWS 기반 클라우드 아키텍처로 설계되어, 웹 요청 처리부터 데이터 분석 및 응답 제공까지의 전체 과정을 자동화하였다. 사용자의 웹 요청은 우선 Amazon Route 53을 통해 도메인 라우팅이 이루어지며, 이후 Application Load Balancer(ALB)로 전달된다. ALB는 요청의 경로(Path)와 트래픽 유형을 분석하여, API 요청은 Spring Boot 서버로, 자연어 처리(NLP) 관련 요청은 FastAPI 서버로 각각 분산시킨다. Spring Boot 서버는 뉴스 데이터 관리, 사용자 인터페이스(UI) 렌더링, API 응답 처리 등의 핵심 비즈니스 로직을 담당한다. 반면 FastAPI 서버는 Python 환경에서 동작하며, 뉴스 크롤링, 형태소 분석, 경제·금융 용어 인식 및 해설 제공 등 텍스트 분석 기능을 수행한다. 두 서버는 모두 Docker 컨테이너 기반으로 배포되어 독립적으로 실행되며, 서로 다른 환경 간의 간섭을 최소화한다. 데이터 저장은 AWS RDS(MySQL)을 통해 이루어지며, 뉴스 본문, 용어 해설, 사용자 학습 기록 등 모든 핵심 데이터가 구조적으로 관리된다. 이를 통해 데이터 접근성과 일관성을 보장하며, 장애 발생 시 자동 백업 및 복구가 가능하도록 설정하였다. 또한 GitHub Actions를 이용한 CI/CD(지속적 통합 및 배포) 파이프라인을 구축하여, 개발자가 코드 변경 사항을 커밋하면 자동으로 빌드, 테스트, 컨

테이너 이미지 생성, 배포가 순차적으로 수행된다. 이를 통해 개발 효율성과 서비스 안정성을 동시에 확보하였다. 보안 측면에서는 AWS IAM(Identity and Access Management) 을 통해 서비스 구성 요소별 접근 권한을 세분화하고, 민감 데이터에 대한 접근을 최소 권한 원칙 (Least Privilege)으로 제한하였다. 또한 ALB와 각 서버 간의 통신은 HTTPS 기반 암호화로 보호되어 데이터 전송 과정의 보안성을 강화하였다. 전체적으로 본 시스템은 확장성(Scalability), 안정성(Stability), 보안성(Security) 을 모두 고려한 모듈형 아키텍처로 구축되었으며, 향후 서버 부하 증가 시 오토 스케일링(Auto Scaling)을 적용하여 유연한 확장이 가능하도록 설계되었다.

## 2) 상세 설계 (기능 설계)

### 1. 뉴스 크롤링

이 시스템은 네이버 경제 뉴스를 대상으로 자동 크롤링을 수행하고, 뉴스 본문에서 등장하는 경제 용어를 추출하여 데이터베이스와 비교한 후 용어 설명과 함께 외부 서버로 전송하는 기능을 제공한다. 이를 통해 경제 뉴스 데이터를 효율적으로 수집하고, 경제 용어 학습 및 분석 자료로 활용할 수 있다. 또한, 수집된 데이터는 경제 동향 분석, 뉴스 요약, 시각화 자료 제작 등 다양한 분야에 활용할 수 있다. 크롤링 프로세스는 크게 네 단계로 구성된다.

```
def get_naver_economy_news_urls_from_list(pages=5):
    """
    네이버 경제 뉴스 목록에서 실제 뉴스 기사 URL을 수집하는 함수
    - pages: 가져올 페이지 수 (한 페이지당 약 20개 기사)
    - 반환: 중복 제거된 뉴스 URL 리스트
    """

    base_url = "https://news.naver.com/main/list.naver"
    headers = {"User-Agent": "Mozilla/5.0"}

    all_links = OrderedDict() # 순서 유지 + 중복 제거

    for page in range(1, pages + 1):
        params = {
            "mode": "LSD",
            "mid": "shm",
            "sid1": "101", # 경제 뉴스 섹션 코드
            "page": str(page)
        }

        response = requests.get(base_url, headers=headers, params=params)
        print(f"📄 {page}페이지 응답 코드: {response.status_code}")

        if response.status_code != 200:
            print(f"❌ {page}페이지 요청 실패")
            continue

        soup = BeautifulSoup(response.text, "html.parser")
        links = soup.select("dt > a")

        for a in links:
            href = a.get("href")
            if href and href.startswith("https://n.news.naver.com"):
                all_links[href] = None # OrderedDict에 key로만 저장 → 중복 제거

    print(f"✅ 총 {len(all_links)}개의 중복 없는 뉴스 링크 수집 완료")
    return list(all_links.keys())
```

첫 번째 단계는 뉴스 URL 수집이다. 사용자가 지정한 페이지 수만큼 네이버 경제 뉴스 목록 페이지를 반복적으로 조회하며, 각 페이지에서 뉴스 기사로 연결되는 링크를 추출한다. HTTP 요청 시 User-Agent를 지정하여 서버 차단을 방지하고, HTML 파싱을 통해 <dt> > <a> 태그에 포함된 링크를 선택한다. 추출된 링크는 OrderedDict에 저장하여 중복을 제거하고 순서를 유지하며, 이 과정에서 https://n.news.naver.com로 시작하는 유효한 뉴스 링크만 수집한다. 페이지 요청이 실패하거나 응답 코드가 200이 아닌 경우에는 해당 페이지를 건너뛰도록 처리하여 안정성을 확보한다. 최종적으로 중복 없는 뉴스 URL 리스트가 반환되며, 이를 기반으로 다음 단계의 처리 대상이 결정된다.

```
def get_news_text(url: str) -> str:
    headers = {
        # 크롤링 시 차단을 피하기 위한 User-Agent 설정
        "User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) "
        "AppleWebKit/537.36 (KHTML, like Gecko) Chrome/91.0.4472.124 Safari/537.36"
    }
    response = requests.get(url, headers=headers) # URL 요청
    if response.status_code != 200:
        print(f"❌ 크롤링 실패: {url} 상태코드 {response.status_code}")
        return "" # 실패 시 빈 문자열 반환

    soup = BeautifulSoup(response.text, "html.parser") # HTML 파싱

    # 네이버 뉴스 본문은 보통 'div#newsct_article' 안에 있음
    article_body = soup.select_one("div#newsct_article")
    if article_body:
        # 본문 텍스트를 줄바꿈 포함해 깔끔하게 추출
        return article_body.get_text(strip=True, separator="\n")
    else:
        print(f"❌ 본문 없음: {url}")
        return "" # 본문 없으면 빈 문자열 반환

# 뉴스 제목과 대표 이미지 URL을 추출하는 함수
def get_title_and_image(url: str):
    response = requests.get(url) # 해당 뉴스 URL 요청
    soup = BeautifulSoup(response.text, "html.parser") # HTML 파싱

    # 제목: h2 태그 중 id가 title_area인 요소의 span에서 텍스트 추출
    title_tag = soup.select_one("h2#title_area span")
    title = title_tag.text.strip() if title_tag else "제목 없음" # 없으면 기본값

    # 대표 이미지: 기사 내 figure 태그 안 img 태그에서 src 속성 추출
    image_tag = soup.select_one("figure img")
    image_url = image_tag["src"] if image_tag and "src" in image_tag.attrs else None # 없으면 None

    return title, image_url
```

두 번째 단계는 뉴스 본문과 메타데이터 추출이다. 수집된 각 뉴스 URL에 대해 HTTP 요청을 수행하고, BeautifulSoup으로 HTML을 파싱한다. 뉴스 본문은 일반적으로 div#newsct\_article 요소 내에 존재하며, 해당 영역의 텍스트를 줄바꿈 포함하여 깔끔하게 추출한다. 본문이 존재하지 않거나 요청이 실패한 경우에는 빈 문자열을 반환하며, 오류 발생 로그를 남겨 추후 확인이 가능하다. 뉴스 제목은 h2#title\_area span 요소에서 추출하며, 대표 이미지는 기사 내 <figure> 태그의 <img> 요소에서 src 속성을 확인하여 가져온다. 이를 통해 뉴스의 핵심 정보와 시각 자료를 함께 확보할 수 있으며, 추후 뉴스 목록 표시나 시각화 자료 제작에 활용할 수 있다.



```

def find_description_from_db(conn, term_input):
    cursor = conn.cursor()
    query = "SELECT * FROM terms WHERE term = ?"
    cursor.execute(query, (term_input,))
    result = cursor.fetchone()
    if result:
        desc1 = result[2] if result[2] else ""
        desc2 = result[3] if result[3] else ""
        desc3 = result[4] if result[4] else ""
        return desc1, desc2, desc3
    else:
        return None, None, None

```

세 번째 단계는 경제 용어 추출 및 설명이다. 추출된 뉴스 본문은 KoNLPy의 Okt 형태소 분석기를 사용하여 명사를 중심으로 단어를 추출한다. 단어 끝에 붙은 조사나 의미 분석에 방해되는 요소는 미리 정의한 조사 목록을 기준으로 제거하며, 추출 과정에서 공백이나 불필요한 기호를 제거한다. 이렇게 처리된 단어들은 등장 순서를 유지하며 중복을 제거한 후, 로컬 SQLite 데이터베이스에 저장된 경제 용어와 비교하여 해당 단어에 대한 설명을 조회한다. 데이터베이스에는 최대 세 가지 설명 항목이 존재하며, 해당 단어가 데이터베이스에 존재하는 경우 각각의 해설을 수집한다. 이를 통해 뉴스 본문에서 등장한 경제 용어에 대한 상세 정보를 구조화된 형태로 확보할 수 있다. 추출된 용어와 설명은 기사와 함께 저장되어, 경제 용어 학습, 뉴스 요약, 혹은 통계 자료 제작 등 다양한 분석에 활용할 수 있다.

네 번째 단계는 외부 서버(Spring)로의 전송이다. 수집된 뉴스 정보와 용어 설명을 JSON 형식으로 구성하여 POST 요청을 수행하며, 전송 상태와 서버 응답 결과를 확인하여 성공 여부를 기록한다. 전송 실패 시에는 예외 처리를 통해 오류 메시지를 기록하고, 재시도 혹은 수동 확인이 가능하도록 처리한다. 이를 통해 수집된 데이터는 중앙 서버에서 통합 관리되며, 후속 분석이나 서비스 제공에 안정적으로 활용될 수 있다.

```

app = FastAPI()

# Spring 서버 주소
SPRING_ENDPOINT = "http://localhost:8080/news/upload"

@app.get("/crawl_auto/")
def crawl_auto(pages: int = Query(5, description="가져올 뉴스 페이지 수")):
    # 1. 자동으로 URL 리스트 수집
    urls = get_naver_economy_news_urls_from_list(pages=pages)

    results = []
    for url in urls:
        title, image_url = get_title_and_image(url)
        content = get_news_text(url)
        terms = extract_and_explain(content, db_path)

```

시스템 전체 구조는 뉴스 수집, 본문/메타데이터 추출, 경제 용어 분석, 외부 전송의 단계별 처리로 구성된다. 각 단계에서는 오류 발생 가능성을 최소화하기 위해 예외 처리를 수행하며, 로그 기록과 상태 확인 기능을 통해 안정성을 확보한다. 이러한 구조를 통해 대량의 경제 뉴스 데이터를 자동으로 수집하고, 뉴스 본문 내 전문 용어를 분석하며, 외부 시스템과 연계하여 활용할 수 있다. 또한, 수집된 뉴스와 용어 데이터를 기반으로 경제 동향 분석, 트렌드 시각화, 용어 학습용 자료 제작, 뉴스 추천 시스템 개발 등 다양한 후속 활용이 가능하다.

## 2. 핵심 키워드 추출 기능 & 오늘의 키워드 기능

핵심 키워드 추출 기능은 각 뉴스 기사에서 의미 있는 경제·금융 관련 단어를 자동으로 추출하고 이를 기반으로 기사별 핵심 키워드와 하루 동안의 핵심 이슈를 알려줄 수 있는 전체 뉴스에 대한 오늘의 키워드 서비스를 제공하는 것이다. 이 기능은 향후 사용자의 관심 뉴스 추천 기능의 기반 데이터로 활용된다.

먼저, 핵심 키워드를 추출하기 위해 크롤링해 온 경제 뉴스 본문을 자동으로 분석하여 해당 기사에서 가장 중요한 단어를 핵심 키워드로 선정한다. 키워드 추출 과정은 크게 형태소 분석 기반 전처리, TF-IDF 기반 중요도 산출, 경제·금융 용어 중심 필터링의 단계로 구성된다.

자연어 처리 기법은 앞선 뉴스 크롤링 단계에서 활용된 것과 동일한 형태소 분석 기반 기술을 그대로 사용하여 적용하였다. 주요 언론사에서 수집된 기사 본문을 형태소 분석(Oktagram)을 통해 조사, 기호 등 불필요한 요소를 제거하고 의미 단위의 명사 및 복합 명사를 추출하는 방식으로 처리된다. 추출된 단어들은 이후 사전에 구축된 용어 집합과 대조되어 유의미한 키워드만을 필터링한다. 이렇게 추출된 키워드는 본문 하이라이팅 기능, 사용자 추천 알고리즘, 오늘의 키워드 등 서비스에서 전반적으로 활용되는 핵심 데이터가 된다.

오늘의 키워드 기능은 하루 동안 적재된 모든 뉴스에서 등장한 키워드를 집계하고 그 빈도수를 기반으로 주요 이슈를 요약해 시각적으로 제시하는 역할을 한다. 오늘의 키워드 탭으로 들어갔을 때 키워드의 빈도수가 높은 순으로 글자 크기를 배치하여 어떤 키워드가 가장 많이 나왔는지 시각적으로 한번에 확인할 수 있고 오른쪽에 배치한 인기 키워드 top10을 이용해 순위별로 주요 키워드를 확인할 수 있다. 시스템은 크롤링해 온 뉴스가 저장될 때마다 키워드 빈도를 갱신하고, 이미 존재하는 키워드는 등장 수를 증가시키며 통계를 관리할 수 있다. 이렇게 관리되는 키워드 정보를 기반으로 생성되는 정보를 기반으로 사용자가 하루의 경제 흐름을 직관적으로 파악할 수 있게 해준다. 특히, 생성된 워드클라우드에서 키워드 크기는 등장 빈도를 직접적으로 반영하기 때문에 사용자는 별도의 분석 과정 없이 당일의 주요 경제 변화를 빠르게 이해할 수 있다. 또한 키워드를 클릭하면 해당 키워드가 포함된 뉴스 목록을 보여주고 뉴스를 클릭했을 때 바로 뉴스 본문으로 이동할 수 있어, 관심 이슈를 중심으로 뉴스를 탐색할 수 있는 탐색 효율을 향상할 수 있다. 데이터 품질 관리 측면에서도 이 기능은 안정적인 운영을 위한 다양한 방어 로직이 필요하다. 뉴스 주소 중복 체크, 용어·키워드 중복 저장 방지, 불완전하거나 의미가 없는 빈 키워드 제거 등 필터링 절차가 포함되어 있으며, 모든 키워드 쓰기 작업은 트랜잭션 내에서 처리되어 배치 작업과 실시간 API 호출이 동시에 이루어져도 데이터 정합성(Data Integrity)이 유지되도록 설계하였다.

### 3. 관심 뉴스 추천 기능

관심 뉴스 추천 기능은 사용자가 읽은 뉴스의 패턴을 학습하여 개개인의 관심사에 따라 맞춤형 뉴스를 자동으로 제시하는 개인화 추천 시스템이다. 사용자가 특정 뉴스를 열람하면 해당 기사에 포함된 경제/금융 키워드가 자동으로 추출되어 user\_news\_log에 저장되고, 이 데이터는 이후 사용자의 관심 키워드를 측정하는 핵심 정보로 활용할 수 있다. 사용자별로 축적된 로그는 특정 키워드에 대한 선호도를 수치화하여 개인화 모델의 입력값이 된다. 추천 시스템은 두 가지의 주요 알고리즘을 결합하여 동작한다.

첫째, TF-IDF(Term Frequency-Inverse Document Frequency) 기반 가중치 계산을 통해 사용자가 자주 접한 키워드를 개인에게 중요한 단어로 판단한다. TF-IDFs는 개별 문서인 뉴스 기사에서 특정 키워드가 얼마나 자주 등장하는지 계산하는 TF와 전체 문서에서 그 단어가 얼마나 희귀한지 계산하는 IDF를 조합하여 단어의 중요도를 산출하는 기법이다. 즉, 뉴스 전체에서 흔히 등장하는 일반적인 단어보다 특정 기사나 특정 사용자 로그에서 빈번하게 등장하는 고유한 단어에 더 높은 가중치를 부여함으로써 사용자별 관심사를 세밀하게 반영할 수 있다. 이 방식은 단순한 빈도 분석보다 개인의 경제적 관심 분야를 명확하게 구분하는 데 유리하며, 특히 금융 도메인처럼 전문 용어가 다양하게 등장하는 문서 환경에 효과적이기 때문에 이 기능을 구현하는 핵심 알고리즘으로 사용했다.

두 번째, Word2vec 기반 의미 벡터 분석을 통해 키워드 간 의미적 유사도를 계산할 수 있다. Word2vec은 단어를 저차원 벡터 공간에 임베딩하여 문맥적으로 유사한 단어들을 서로 가깝게 배치한다. 이를 활용하면 시스템은 직접적으로 연관되지 않은 관련 이슈까지 함께 추천할 수 있다. 이 의미 기반 확장은 TF-IDF만으로 포착하기 어려운 잠재적 관심사를 반영하여 더 정교한 개인화 추천 시스템을 구현하는 데 활용할 수 있다.

추천 과정은 사용자가 메인화면에 재접속할 때 실시간으로 수행된다. 백엔드는 먼저 사용자의 관심 카테고리인 user\_interest\_category를 확인한 뒤, 해당 카테고리 및 연관된 키워드를 news\_keyword 데이터와 비교하여 후보 기사를 선별한다. 이후 TF-IDF 가중치와 Word2Vec 유사도 점수를 조합해 최종 추천 순위를 산출한다. 최근 전체 사용자에게 높은 관심을 받은 트렌드 키워드(keyword\_trend)까지 함께 반영하여 개인 취향과 사회적 이슈가 균형 있게 드러나는 추천 결과를 제공한다.

#### 4. 단답형 퀴즈 설계

##### - 설계 흐름

사용자 요청(/api/quiz/short-answer) → 사용자 단어장 조회 → 단어장 비어있음 확인 → Quiz 엔티티 생성 → 각 단어에 대해 Glossary 확인 및 문제 문장 생성 → 정답 단어 □ 치환 → 초성 힌트 생성 → QuizTerm 생성 및 저장 → QuizDto 구성 → 클라이언트로 결과 반환

##### - 제출(채점) 흐름

사용자 정답 제출 → Term 조회 → 입력값 정규화 → 정답과 비교 → 난이도 기반 점수 계산 → 정답 단어 단어장에서 제거 → UserPoint 업데이트 → 배지 평가 및 부여 → 점수·총 포인트·정답 여부 반환

단답형 퀴즈 기능은 사용자가 단어장에 저장한 경제·금융 용어를 기반으로 문제를 자동 생성하여 학습 효율을 높이는 기능으로 설계되었다. 시스템은 Controller, Service, Repository, Domain으로 구성된 계층형 아키텍처를 기반으로 하며, REST API를 통해 문제 생성부터 채점까지의 모든 흐름을 처리한다. QuizController는 퀴즈 관련 요청을 수신하는 역할을 수행하며, QuizService는 문제 출제, 초성 힌트 생성, 해설 기반의 문제 문장 구성, 사용자 답변 채점, 점수 계산, 단어장 업데이트, 포인트 및 배지 지급 등 핵심 비즈니스 로직을 담당한다. Repository 계층은 Quiz, QuizTerm, QuizResult, UserVocabulary 등의 엔티티에 대한 데이터베이스 접근을 담당하며, 도메인 엔티티는 퀴즈 기능을 구성하는 데이터 구조의 기반이 된다. 전체 구조는 타입 기반 설계를 바탕으로 단답형뿐 아니라 가로세로 퍼즐 등 다양한 퀴즈 타입 확장이 가능하도록 설계되었다.

데이터 모델은 사용자(User)를 중심으로 단어장(UserVocabulary), 용어(Term), 해설(Glossary)와 퀴즈(Quiz)·퀴즈 문제 항목(QuizTerm)·결과(QuizResult)가 연결되는 구조로 구성된다. Quiz는 생성 시점, 퀴즈 타입, 문제 목록과 결과 목록을 포함하며, QuizTerm은 출제된 용어와 초성 힌트를 포함한다. 용어(Term)는 빈도값(frequency)을 가지고 있어 난이도 계산에 활용되고, Glossary는 용어에 대한 짧은 해설을 여러 개 포함하여 문제 문장 생성에 사용된다. 이러한 구조는 출제, 채점, 통계 등 다양한 기능 확장에 용이하도록 고려되었다.

단답형 퀴즈는 사용자의 단어장 기반으로 자동 생성되며, 단어장의 모든 용어에 대해 해설이 존재하는 경우에만 QuizTerm을 구성한다. 문제 문장은 Glossary의 shortDefinition에서 정답 단어를 '□' 문자로 치환하여 구성되며, Glossary가 여러 개일 경우 최대 3개의 문장을 번호와 함께 제공하여 학습적 효과를 높인다. 초성 힌트는 한글 유니코드 구조를 활용하여 초성을 계산하는 방식으로 생성되며, 이는 사용자가 정답을 유추하는 데 중요한 힌트로 작용한다. 문제 생성 단계에서는 단어장 조회 → Quiz 엔티티 생성 → 용어별 QuizTerm 생성 → 초성 힌트 적용 → QuizTerm 저장 → QuizDto 생성의 순서로 처리된다.

사용자가 제출한 답변은 정규화 과정을 거친 뒤 정답과 비교된다. 정규화는 특수문자 제거, 공백 제거, 영문 소문자 변환을 포함하여 사용자의 입력 오류나 표현 차이가 결과에 영향을 주지 않도록 설계되었다. 정답 여부는 정규화된 사용자 답변과 정답 용어의 정규화된 형태를 비교하여 판단하며, 정답일 경우 난이도 값을 기반으로 점수가 계산된다. 난이도는 해당 용어의 frequency 값을 그대로 활용하며, 기본 점수는 난이도 × 10점이다. 점수 계산 후 정답인 용어는 자동으로 단어장에서 제거되어 이후 퀴즈에서 다시 등장하지 않도록 함으로써 학습 진도를 자연스럽게 관리할 수 있다.

퀴즈 결과는 QuizResult 엔티티에 기록되며, 사용자의 포인트 시스템과 연동된다. 정답으로

획득한 점수는 UserPoint에 누적되며, 포인트 증가에 따라 BadgeGrantService를 통해 자동으로 배지가 평가되고 부여된다. 이를 통해 사용자는 학습 과정에서 지속적인 보상을 경험하면서 지속적인 퀴즈 참여를 유도하게 된다. 또한 정답/오답 여부는 프론트엔드로 즉시 전달되며, 오답일 경우 정답 단어를 함께 제공하여 학습 효과를 극대화할 수 있도록 구성되어 있다. API는 퀴즈 생성 API와 제출 API 두 가지 중심으로 구성된다. 퀴즈 생성 API는 GET /api/quiz/short-answer로 호출되며, 세션 기반 인증을 통해 사용자 정보를 확인한 뒤 답변 항목 리스트를 포함한 QuizDto를 반환한다. 제출 API는 POST /api/quiz/submit으로 구현되며, quizId와 answers 맵을 전달받아 채점 및 점수 계산, 단어장 업데이트, 포인트 지급, 배지 지급을 처리한 뒤 score·takenAt·정답 여부 목록·총 포인트 등을 포함한 응답을 반환한다. 에러 처리 또한 사용자 인증 실패, 엔티티 조회 실패, 단어장 데이터 없음 등의 상황을 고려하여 설계되었다.

전체 기능은 성능과 확장성을 고려하여 구현되었다. 한글 초성 추출은 정규화된 유니코드 계산을 통해 경량화된 방식으로 수행되며, 데이터 접근은 지연 로딩(LAZY)과 saveAll을 통한 배치 저장을 활용해 효율성을 높였다. 문제 문장 생성에서는 정규 표현식을 통해 빠른 문자열 변환을 수행하며, 정답 치환, 다중 Glossary 처리 등도 동적으로 구성된다. 또한 초성 힌트 추출과 입력 정규화는 재사용 가능한 Utility로 분리하여 다른 퀴즈 타입에서도 동일하게 활용할 수 있도록 설계했다.

향후 개선 방향으로 사용자 정답 패턴 기반의 난이도 자동 조정, 틀린 문제 중심의 복습 모드, 학습 통계 기능, 제한 시간 기능, 단계별 힌트 기능 등이 고려될 수 있으며, 현재 구조는 이러한 확장을 무리 없이 수용할 수 있도록 설계되어 있다. 결론적으로 단답형 퀴즈 기능은 사용자 개인화 기반의 문제 출제, 학습 진도 추적, 포인트 및 배지 시스템과의 연동을 통해 높은 학습 동기와 반복 학습을 유도하는 구조로 구축되었으며, 모듈화된 아키텍처를 기반으로 다양한 퀴즈 유형으로 확장 가능한 기능적·구조적 장점을 보유하고 있다.

## 5. 가로세로 낱말 퀴즈 설계

### - 가로세로 낱말 퀴즈 기능 설계 흐름

사용자 요청(/api/quiz/crossword) → 최근 단어장 5개 조회 → Glossary 있는 단어만 필터 → 각 가로 단어별 세로 단어 후보 탐색(공통 글자 기반) → 21×21 보드 초기화 → 가로·세로 단어 세트 구성 → 겹침 위치 계산 → 보드 경계 체크 → 기존 배치 충돌 검사 → 위치 재시도 및 배치 확정 → CrosswordItem 변환 → 초성 힌트 생성 → QuizDto 구성 → 클라이언트 응답 반환

### - 제출(채점) 흐름

사용자 정답 제출 → Term 조회 → 입력 값 정규화 → 정답과 비교 → 난이도 기반 점수 계산 → 정답 단어 단어장에서 제거 → UserPoint 업데이트 → 배지 평가 및 부여 → 점수·총 포인트·정답 여부 반환

가로세로 낱말 퀴즈 기능은 사용자가 최근에 단어장에 추가한 경제·금융 용어를 기반으로 전통적인 십자말풀이 형식의 퍼즐을 자동 생성하는 기능으로, 21×21 크기의 보드에 가로 단어와 세로 단어가 교차하도록 배치하여 용어 간 연관성을 자연스럽게 학습하도록 설계되었다. 시스템은 Controller, Service, Repository, Domain Layer로 구성된 계층형 아키텍처를 기반으로 구현되며, QuizController가 /api/quiz/crossword 및 /api/quiz/submit-crossword API를 제공하고, QuizService가 퀴즈 생성·배치·채점·포인트 처리 등의 핵심 로직을 수행한다. 또

한 단어 배치 과정에서 필요한 정보를 관리하기 위해 내부 클래스인 PlacedWord와 DTO 변환을 위한 CrosswordItem을 활용하여 데이터 구조를 분리·관리하고 있다.

퀴즈 생성 과정은 사용자 단어장에서 최근 추가된 5개 단어를 조회하는 것으로 시작된다. 각 단어는 가로 단어로 사용되며, 이와 겹치는 글자를 가진 세로 단어를 전역 경제·금융 용어 목록에서 찾아 후보군을 구성한다. 3글자 이상인 단어는 첫 번째 글자와 세 번째 글자를 기준으로 2개의 세로 단어를 선택하며, 2글자인 단어는 첫 번째 글자를 기준으로 1개의 세로 단어만 선택한다. 선택된 단어는 Glossary가 존재하고 이미 사용되지 않은 단어만 포함되도록 필터링된다. 이후 21×21 보드를 초기화하고, 가로 단어와 세로 단어를 하나의 세트로 묶어 연속적으로 배치한다. 보드 배치 시 가로 단어와 세로 단어의 겹침 위치를 계산하고, 배치 위치가 보드 범위를 벗어나지 않는지, 이미 배치된 세트와 충돌하지 않는지 확인한 뒤 배치를 수행한다. 충돌이 있는 경우 다른 위치로 이동해 재시도하며, 최대 100회 범위 내에서 배치가 실패하면 해당 세트는 출력 대상에서 제외된다. 이러한 방식은 보드 내 단어 배치의 안정성을 확보하고 불필요한 무한 루프를 방지하기 위해 설계되었다.

배치된 단어는 CrosswordItem으로 변환되며, 각 단어는 row·col·length·direction·문제번호·정답단어·해설·초성 힌트를 포함하는 구조로 구성된다. 해설은 Glossary의 shortDefinition을 사용하고, 초성 힌트는 한글 초성 추출 알고리즘을 통해 계산하여 사용자가 정답을 추론하는데 도움을 주도록 한다. 이후 모든 항목은 QuizDto 형태로 반환되며, 클라이언트는 이를 바탕으로 가로세로 퍼즐 UI를 구성한다.

채점 과정은 제출된 단어별 사용자 답변을 정규화하여 정답과 비교하는 방식으로 처리된다. 정규화 규칙은 단답형 퀴즈와 동일하게 특수문자 제거, 공백 제거, 영문 소문자 통일을 포함한다. 정답이 맞으면 해당 Term의 frequency 값을 기반으로 난이도 점수를 계산하며, 각 정답 단어는 사용자 단어장에서 자동 제거되어 학습 진도 관리를 돕는다. 채점이 완료되면 점수는 UserPoint 엔티티에 기록되며 누적 포인트가 갱신된다. 이어서 BadgeGrantService에서 사용자의 누적 포인트를 기준으로 신규 배지를 평가·부여한다. 최종적으로 점수, 누적 포인트, 단어별 정답 여부가 클라이언트로 반환된다.

핵심 설계의 특징은 크게 네 가지로 요약된다. 첫째, 지능형 단어 배치 알고리즘을 통해 가로 단어와 세로 단어 간의 교차점을 자동 분석하고, 보드 영역 충돌을 방지하는 구조를 구현하였다. 이는 단어 간 공통 글자 탐지, 세트 단위 배치, 경계 오류 방지, 중복 배치 방지 등을 포함하여 안정적인 퍼즐을 구성하는 데 중점을 둔다. 둘째, 사용자 맞춤형 학습 경험을 제공하기 위해 최근 학습한 단어를 중심으로 가로 단어를 구성하고, 연관 단어를 세로 단어로 출제하여 학습 맥락을 확장한다. 또한 정답 단어 자동 제거 기능을 통해 효과적인 반복 학습을 유도한다. 셋째, 확장 가능한 아키텍처를 기반으로 가로 단어와 세로 단어를 세트 단위로 처리하여 동적 크기 조정, 다양한 겹침 패턴, 향후 복잡한 퍼즐 생성 알고리즘 확장에도 대응할 수 있는 구조를 확보했다. 넷째, 사용자 경험을 고려한 설계로 시각적 직관성이 높은 십자말 풀이 UI 제공, 초성 힌트 지원, 즉시 정답 피드백을 통해 사용자의 참여도를 높이는 방향으로 구성되어 있다.

기술적으로는 공통 글자 탐지 알고리즘, 최대 시도 제한 기반 충돌 회피 전략, 세트 간 여백 부여를 통한 가독성 향상, FetchType.LAZY 및 스트림 API 활용을 통한 성능 최적화가 포함된다. 또한 예외 처리 측면에서는 단어장 비어 있음, Glossary 미존재, 겹침 글자 미발견, 보드 경계 초과, 최대 시도 초과 등의 오류를 감지하여 안정적인 실행을 보장하도록 구현되어 있다.

## 6. Spring Batch 기반 뉴스 처리 파이프라인

뉴스 적재 파이프라인은 Python 기반 크롤러가 전달하는 뉴스 데이터 묶음을 안정적으로 받아들이고, 이를 정제한 뒤 중복을 제거하고, 서비스 전용 데이터베이스에 저장하고, 키워드 연동 정보까지 함께 반영하는 전자동 데이터 처리 흐름을 구축하는 것을 목표로 한다. 이 과정 전체는 세 단계의 정형화된 배치 흐름으로 구성되며, 최종적으로 생성되는 데이터는 홈 화면, 키워드 클라우드, 단어장 기능 등 서비스 전반의 모든 기능에서 공통 기반으로 활용된다. 따라서 이 파이프라인은 단순 적재 작업을 넘어, 플랫폼 전체의 품질과 완성도를 좌우하는 핵심 요소라고 할 수 있다.

전체 배치는 수집->정제->처리 라는 순차 흐름으로 구성되며, 각 단계는 독립적인 처리 단위를 갖지만 내부적으로는 메모리에 올려둔 임시 저장소를 통해 자연스럽게 연결된다. 여러 단계로 나누어 처리하는 이유는 초기 원본 데이터를 최대한 변형 없이 수집한 뒤, 정제와 중복 처리에서 발생할 수 있는 오류를 빠르게 감지하고, 마지막 저장 과정에서는 최대 성능으로 대량 데이터를 처리하기 위함이다.

첫 번째 단계인 "수집 단계"는 매우 큰 단위를 다루기 때문에 한 번의 외부 요청으로 전체 데이터를 수집하도록 설계하였다. 실행 시점마다 설정된 주소로 POST 요청을 보내면 Python 크롤러가 일정 기간 동안 수집한 뉴스 배열 전체를 반환하며, 배치 시스템은 이를 단 한 번만 수신한다. 이 과정에서 통신 지연이나 연결 오류를 예방하기 위해 요청-응답 제한 시간을 별도로 지정해 두었다. 수신된 뉴스 데이터는 즉시 임시 저장소에 보관되며, 이후 단계들이 이를 반복적으로 읽어 사용할 수 있도록 "한 번 받고 여러 번 읽는 구조"를 채택했다. 이 방식을 사용함으로써 외부 서버를 불필요하게 반복 호출하거나 네트워크 비용을 낭비하지 않게 된다.

두 번째 단계인 "정제 단계"는 전체 파이프라인에서 가장 많은 판단을 수행하는 구간이다. 임시 저장소에 쌓인 원본 데이터를 하나씩 읽어 제목, 본문, 이미지, 출처, 발행 시각 등 필요한 정보를 뽑아내고 서비스 내부의 뉴스 구조로 변환한다. 이 과정에서 가장 중요한 요소는 "중복 제거"이다. 같은 뉴스가 여러 번 수신될 가능성을 고려해 우선적으로 중복 여부를 확인하고, 동일한 주소를 가진 뉴스는 바로 제외함으로써 불필요한 처리 비용과 저장 공간 낭비를 막는다. 또한 언론사 정보가 누락된 경우처럼 데이터 품질이 흔들릴 수 있는 상황에서는 기본값을 자동으로 채워넣어 후속 기능들이 일관된 정보를 기반으로 동작하도록 했다. 정제가 끝난 뉴스는 다시 별도의 임시 저장소에 보관되며, 다음 단계에서 실제 저장 작업을 수행할 수 있도록 준비된다.

세 번째 단계인 "저장 단계"에서는 정제된 뉴스 데이터가 최종적으로 데이터베이스에 반영된다. 이 단계에서는 다시 한 번 중복 여부를 확인해 잠재적인 오류를 막고, 중복이 아닌 뉴스만 저장한다. 특히 뉴스에 포함된 키워드 정보는 뉴스 본문과 분리된 형태로 별도의 테이블에 저장되며, 뉴스와 키워드 간의 관계가 명확히 연결되도록 설계했다. 배치 처리 중에는 키워드가 임시 형태로 뉴스 객체에만 보관되지만, 실제 저장 시점에 이 정보가 키워드 테이블의 개별 레코드로 풀려 들어가며, 뉴스와 키워드 간의 연관 정보가 형성된다. 이를 통해 서비스는 각 뉴스에 어떤 키워드가 연결되어 있는지 빠르게 조회할 수 있고, 키워드 클라우드나 개인화 추천과 같은 기능에서 이를 적극적으로 활용할 수 있다.

이 파이프라인은 정해진 주기마다 자동으로 실행되며, 주기는 설정 파일에서 2분 간격으로 조정되어 있다. 실행 시마다 현재 시각을 실행 파라미터로 기록해두어 동일 작업을 반복 실행하거나 재시작할 필요가 있을 때 기준점으로 활용할 수 있도록 구성하였다. 모든 실행 정보와

결과는 로그로 남기고 있으며, 배치 처리 과정에서 수집되는 메타 정보는 배치 전용 시스템 테이블에 자동 저장된다. 이를 통해 실행 성공·실패 내역, 소요 시간, 처리 건수 등을 투명하게 조회할 수 있고, 예상치 못한 오류 발생 시 문제 원인을 빠르게 추적할 수 있다.

데이터 구조 관점에서는 뉴스와 키워드의 구조가 명확하게 분리되어 관리된다. 뉴스는 제목, 본문, 언론사, 발행일, 주소를 중심으로 한 단일 구조이며, 각 뉴스는 여러 개의 키워드를 가질 수 있는 1:N 형태의 관계를 갖는다. 배치 중에만 존재하는 임시 키워드 정보는 실제 저장 단계에서 완전한 키워드 엔티티로 변환되어 뉴스와 연결된다. 이렇게 저장된 정보는 홈 화면의 최신, 맞춤 뉴스 구성, 키워드 클라우드 계산, 단어장 기능 등에서 다양하게 사용된다.

오류와 복구 전략 역시 자연스럽게 파이프라인에 녹아 있다. 수집 단계와 저장 단계 모두에서 중복 검증을 진행해 잘못된 데이터가 유입되는 것을 차단하며, 외부 통신 실패는 즉시 배치 실패로 처리해 운영자가 바로 감지할 수 있도록 한다. 또한 임시 저장소는 비우기 기능을 제공해, 여러 작업 실행 간 데이터가 섞이는 문제를 방지할 수 있도록 되어 있다. 현재 구조는 단일 배치 인스턴스를 기준으로 안정적으로 동작하도록 설계되어 있지만, 필요한 경우 작업 종료 시 임시 저장소를 자동으로 비우는 흐름을 추가하는 개선도 고려할 수 있다.

## 7. NewsService 기능설계

NewsService는 뉴스 도메인을 총괄하는 핵심 비즈니스 계층으로, 뉴스 저장, 조회, 용어 하이라이팅, 키워드 통계, 사용자 행태 분석 등 다양한 기능을 하나의 일관된 구조 안에서 처리한다. 이 계층은 여러 저장소(뉴스, 경제·금융 용어, 용어집, 사용자 로그, 키워드 통계 등)를 동시에 다루는 만큼 데이터 정합성과 성능 최적화가 중요한 역할을 담당한다. 전체 클래스는 기본적으로 트랜잭션 경계를 적용해 쓰기 작업의 일관성을 보장하고, 조회 작업에는 읽기 전용 옵션을 부여해 효율적인 데이터 접근이 가능하도록 구성돼 있다.

가장 기본적인 역할은 뉴스 저장이다. 전달된 뉴스가 이미 저장된 주소를 가진 경우 같은 항목을 다시 저장하지 않고 기존 식별자를 곧바로 반환하여 멱등성을 보장한다. 이는 크롤러, 배치, 수동 입력 등 여러 경로에서 같은 뉴스가 반복 수신될 수 있는 상황을 대비한 구조다. 반대로 신규 뉴스라면 저장 후 생성된 식별자를 반환해 후속 로직에서 활용하게 한다. 뉴스 저장과 함께 키워드가 제공되는 경우에는 연동 작업도 함께 처리한다. 먼저 뉴스 저장을 통해 ID를 확보한 뒤, 키워드 목록을 하나하나 키워드 엔티티로 변환하고 뉴스와 연결하여 저장한다. 동시에 키워드 사용 빈도를 증가시키는 통계 작업이 이루어지는데, 이는 키워드 클라우드나 트렌드 분석에서 직접 활용되는 중요한 기초 데이터가 된다. 이러한 흐름은 배치 파이프라인과 동일한 정책을 공유하기 때문에 수동 업로드나 보정 작업에도 같은 도메인 규칙을 재사용할 수 있다.

조회 기능은 목록, 페이징, 단건조회 등 기본적인 형태를 모두 지원하며, 홈 화면과 상세 페이지에서 공통적으로 사용한다. 특히 뉴스 본문을 경제·금융 용어 기반으로 하이라이트하는 기능은 플랫폼만의 차별적 사용자 경험을 제공한다. 이 기능은 데이터베이스에 저장된 경제·금융 용어 목록을 기반으로 본문 내 용어를 찾아내고, 각 용어 주변의 조사까지 인식해 올바르게 감싸도록 구성돼 있다. 이미 하이라이트 처리가 된 본문이 들어오는 경우에도 중복 <mark>가 생기지 않도록 기존 태그를 제거한 뒤 새롭게 적용한다. 이 기능은 뉴스에서 전문 용어를 빠르게 파악하고 학습할 수 있도록 돕는 핵심 서비스 요소다.

경제·금융 용어 및 용어집 저장 기능도 중요한 역할을 맡고 있다. Python 서버가 전달하는 용어 배열을 입력받아 용어 테이블과 용어집 테이블에 차례로 기록하는데, 이미 존재하는 용어



인지 여부를 우선 판별한 뒤 신규 용어만 생성한다. 용어에 연결되는 간단한 정의 역시 중복 여부를 검사한 뒤 없는 항목만 추가한다. 이를 통해 전체 서비스에서 사용되는 용어, 정의 데이터의 품질을 보장하고, 하이라이팅 기능, 단어장 기능, 검색 기능 등 다양한 기능의 기반으로 활용할 수 있게 된다.

사용자 행태 분석 기능도 NewsService의 중요한 영역 중 하나다. 사용자가 어떤 뉴스를 얼마나 자주 클릭했는지를 기록해두었다가, 특정 사용자가 많이 본 뉴스 순으로 정렬해 개인화된 추천 데이터를 구성할 수 있다. 이 데이터는 홈 화면 개인화 영역이나 향후 추천 알고리즘에서 직접적으로 활용될 수 있다. 사용자의 클릭 로그는 뉴스 상세 보기 진입 시마다 기록되며, 사용자, 뉴스, 열람 시각이 함께 저장된다. 이를 통해 개인화 추천뿐만 아니라 활동 통계, 배지 시스템, 이용 패턴 분석 등 다양한 기능으로 확장할 수 있는 여지가 생긴다.

키워드 통계 기능은 전체 뉴스에서 어떤 키워드가 얼마나 쓰였는지를 관리한다. 키워드가 새롭게 등장하면 빈도를 증가시키고, 필요시 감소시키는 작업도 가능하다. 이를 기반으로 상위 키워드를 조회해 트렌드 클라운드를 구성하거나 특정 뉴스에서 등장한 키워드를 우선 순위로 정렬해 화면에 제공한다. 또한 키워드를 기반으로 뉴스 목록을 검색해 사용자 요청에 즉시 응답할 수 있도록 조회 결과를 화면용 데이터로 변환하는 역할도 맡고 있다.

운영 관점에서는 모든 쓰기 작업이 트랜잭션 범위 안에서 처리되도록 구성해 배치와 API가 동시에 호출되더라도 데이터 정합성을 유지한다. 뉴스 주소 중복 체크, 용어집 중복 방지, 빈 키워드 필터링 등 방어 로직이 곳곳에 포함돼 있어 입력 데이터 품질이 일정 수준 이하라도 서비스가 안정적으로 동작하도록 했다. 사용자 클릭 기록 등 주요 이벤트는 로그로 남겨두어 서비스 운영 시 모니터링이나 문제 분석에 활용할 수 있다.

마지막으로, 향후 개선 여지는 하이라이팅 기능의 성능 최적화, 키워드 통계나 사용자 로그를 이벤트 기반 시스템으로 분리해 확장성을 확보하는 방향, 그리고 뉴스 저장 과정에서 엔티티 재조회 없이 영속 객체를 재활용할 수 있는 구조 개선 등이 있다. 이러한 개선사항은 트래픽 증가나 데이터 규모 확대에 대비해 시스템 성능과 운영 효율성을 더욱 높여줄 수 있다.

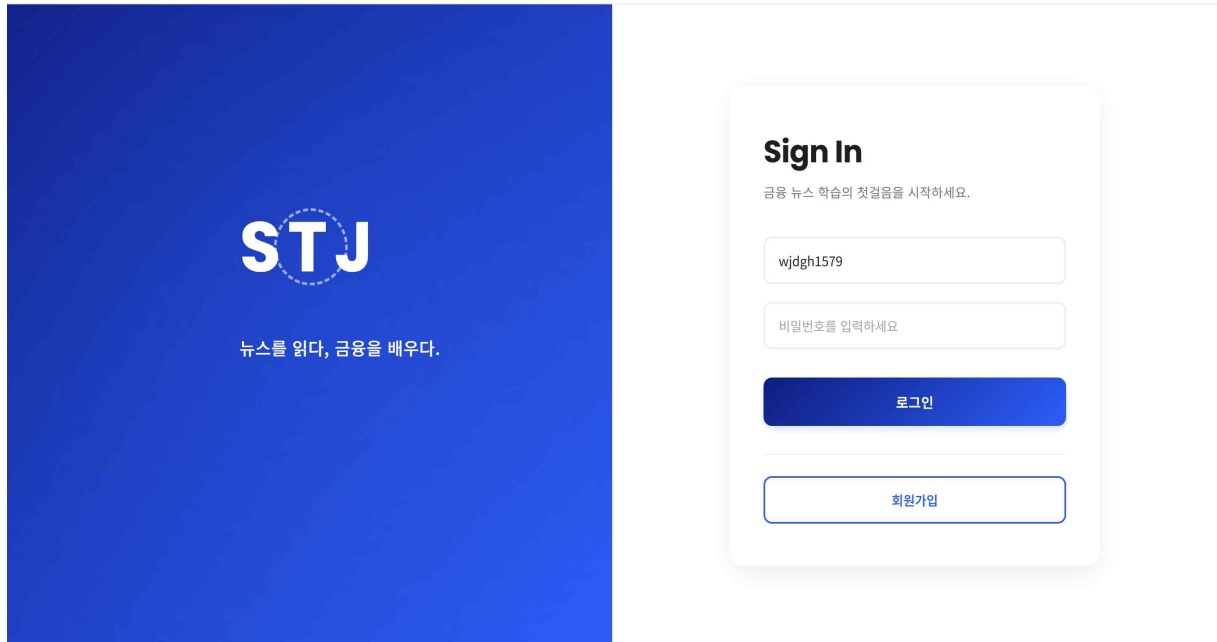
## 나. 수행 방법 및 추진 과정

성명	역할
염정호	AI, NLP활용해 단어 추출, 분류, NER 학습 모델 개발
김준원	AI, NLP활용해 단어 추출, 분류, NER 학습 모델 개발
박세영	백엔드, 서버 구축, 뉴스 데이터 처리, AI 연동, API 자동화
최현서	데이터베이스 구축, 뉴스, 단어 API 및 데이터 처리

일정 (월차)	내용	세부 내용	비고
1	프로젝트 기획	문제 정의 및 인식	
		프로젝트 기획 및 시장조사	
		기능 구체화 및 기술 스택 선정	
2	DB 구축 및 백엔드 설계	프레임 워크 선정, 기본환경 세팅	
		경제·금융 용어 DB스키마 설계 및 구축	
		API구현	
		API테스트 및 개선	
3	NPL기반 용어 해설 기능 제작	뉴스 크롤링 기본 구조 제작	
		데이터 전처리 및 DB저장 구현	
		API테스트 및 크롤링 개선	
4	요약 및 해설 기능 제작	요약 및 해설 기능 제작	
5	추가기능 구현	퀴즈 및 단어장 기능 구현	
6	프론트 엔드 구축	UI설계 및 인터페이스 구축	
		백엔드 - 프론트엔드 연동	
7	통합테스트 진행	통합 테스트 및 오류 수정	
		발표 준비 및 마감	

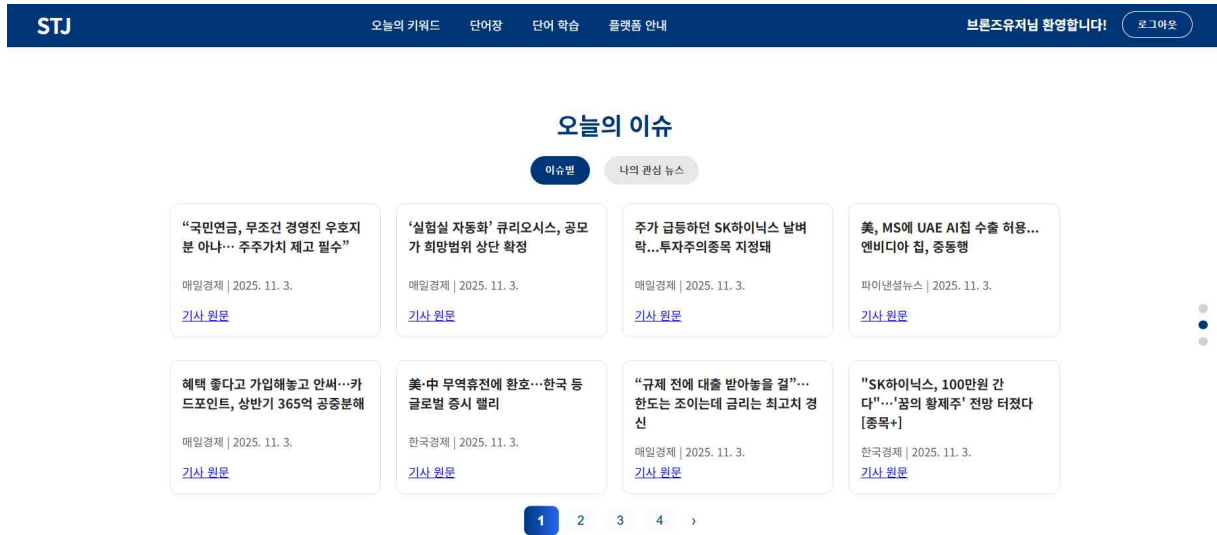
## 다. 최종 결과물

### 1. 로그인 화면



이 화면은 사용자가 서비스를 이용하기 위해 계정 인증을 진행하는 로그인 기능을 제공한다. 사용자는 아이디와 비밀번호를 입력해 본인임을 확인한 뒤, 개인화된 기능(단어장, 학습 기록, 뉴스 추천 등)에 접근할 수 있다. 로그인 정보가 올바를 경우 즉시 메인 페이지로 이동하며, 잘못된 경우 오류 메시지를 통해 재입력을 안내한다. 또한 계정이 없는 사용자를 위해 회원가입 기능을 함께 제공하여 신규 이용자가 쉽게 서비스를 시작할 수 있도록 구성하였다.

## 2. 메인화면



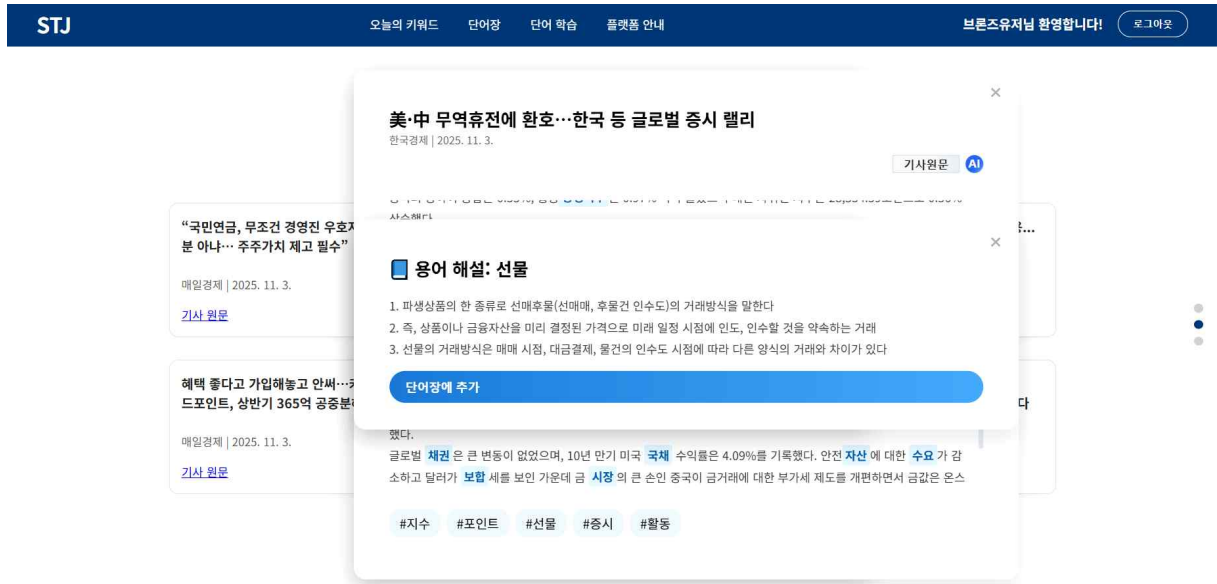
로그인 후 사용자는 첫 화면에서 최신 경제 뉴스를 한눈에 확인할 수 있다.

해당 화면의 상단에는 '오늘의 키워드', '단어', '단어 학습', '플랫폼 안내' 등의 주요 메뉴가 배치되어 있어 사용자가 원하는 기능으로 쉽게 이동할 수 있도록 구성되어 있다.

화면 중앙의 '오늘의 이슈' 섹션에서는 최근 업로드된 경제 뉴스 기사들이 카드 형태로 정리되어 제공된다. 각 뉴스 카드는 제목, 출처, 게시일 정보를 포함하며, 사용자가 '기사 원문' 버튼을 클릭하면 해당 기사의 상세 내용을 확인할 수 있다.

또한 하단의 페이지네이션 기능을 통해 사용자는 여러 페이지에 걸쳐 더 많은 뉴스를 탐색할 수 있다. 이를 통해 사용자는 로그인 직후 실시간 경제 이슈를 빠르게 파악하고, 관심 있는 기사를 선택하여 읽을 수 있는 구조로 설계되었다.

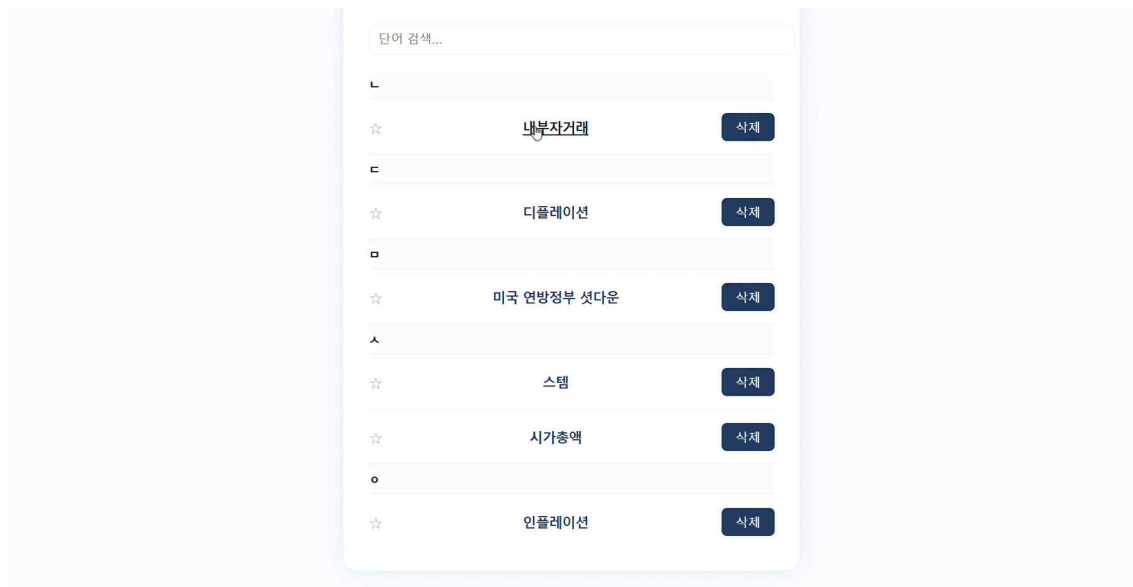
### 3. 경제·금융 용어 해설기능



사용자가 '오늘의 이슈' 화면에서 원하는 기사를 클릭하면, 팝업 형태로 기사 전문이 표시된다.

이때, 본문 내에 금융·경제 관련 주요 용어가 자동으로 하이라이트 처리되어 사용자가 한눈에 확인할 수 있도록 구성되어 있다.

하이라이트된 단어를 클릭하면, 데이터베이스(DB)에 사전에 등록된 해당 용어의 정의 및 해설이 화면 하단에 제공된다. 예를 들어 '선물'과 같은 용어를 클릭하면 그 의미와 관련 설명이 즉시 표시되어, 사용자가 기사를 읽는 흐름을 방해받지 않고 내용을 이해할 수 있다.



또한, 사용자는 '단어에 추가' 버튼을 클릭해 해당 용어를 개인 단어장에 저장할 수 있으며, 이후 학습 메뉴에서 반복적으로 복습할 수 있다.

이 기능은 뉴스 콘텐츠 소비를 단순한 정보 습득이 아니라 금융 문해력 향상 학습 과정으로 확장시키는 역할을 한다.

## 4. AI 뉴스 요약 제공



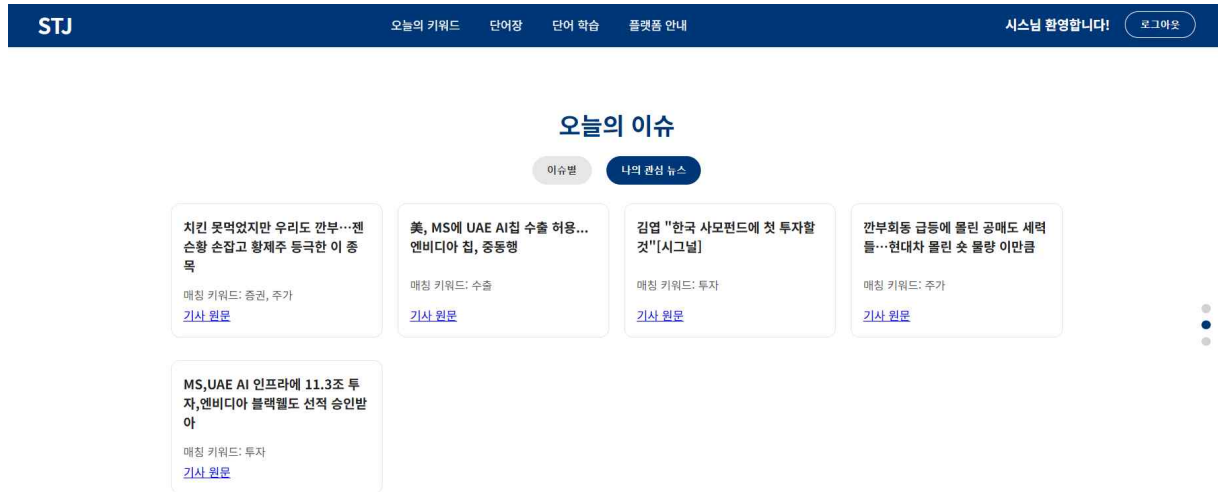
본 시스템은 사용자가 기사 본문을 모두 읽지 않아도 핵심 내용을 파악할 수 있도록 AI 기반 뉴스 요약 기능을 제공한다.

뉴스 상세 보기 팝업 화면 우측 상단에 위치한 “AI 요약보기” 버튼을 클릭하면, 인공지능이 기사 전체 텍스트를 분석하여 중요 문장과 핵심 정보를 자동으로 정리한 요약문을 생성한다.

요약 방식은 단순 문장 축약이 아닌, 기사에서 중요한 주제·사건·수치를 중심으로 내용을 선별하도록 설계되었다. 또한 의미 기반 자연어 처리 모델을 적용하여 기사의 흐름과 맥락을 유지하면서 핵심 메시지를 전달하도록 구성하였다.

해당 기능을 통해 사용자는 긴 기사 내용을 읽는 데 걸리는 시간을 단축할 수 있으며, 특히 경제·정책 등 전문 용어가 포함된 기사에서도 주요 내용을 쉽게 이해할 수 있다. 이를 통해 뉴스 소비의 효율성을 높이고 학습 플랫폼으로서의 활용성을 강화하였다.

## 5. 사용자 키워드 기반 관심 뉴스추천



본 플랫폼은 사용자의 뉴스 소비 패턴을 분석하여, 개인에게 맞춤형 뉴스를 우선적으로 보여주는 키워드 기반 추천 알고리즘을 적용하였다.

뉴스를 수집하는 과정에서 시스템은 기사 본문에 포함된 단어를 형태소 분석을 통해 분리하고, 그중 경제·금융 관련 핵심 단어를 자동으로 키워드로 추출한다.

사용자가 웹페이지에서 특정 뉴스를 열람하게 되면, 해당 기사에 포함된 키워드가 사용자 로그 데이터에 기록된다. 이를 통해 시스템은 사용자가 어떤 키워드에 관심을 갖고 있는지 지속적으로 파악하며, 한 명의 사용자에게 대해 관심 키워드 데이터가 누적된다.

이후 사용자가 다시 뉴스 화면에 접근하면, 저장된 로그 데이터를 기반으로 TF-IDF 유사도 분석 및 Word2Vec 벡터 연산을 수행하여 키워드 간 연관도를 계산한다. 연관도가 높은 키워드를 포함하는 기사는 자동으로 리스트 상단에 배치되며, 사용자는 별도의 조작 없이도 자신의 관심 분야와 관련된 뉴스를 우선적으로 확인할 수 있다.

이 기능을 통해 사용자는 단순히 뉴스 소비자가 아니라, 자신의 관심사에 따라 정보가 큐레이션되는 경험을 하게 되며, 플랫폼은 학습 참여도와 정보 탐색 효율성을 함께 높일 수 있다.

## 6. 오늘의 키워드



본 시스템은 하루 동안 수집된 경제 뉴스 기사에서 핵심 키워드를 자동으로 추출하여, 등장 빈도에 따라 오늘의 키워드로 분류한다. 추출된 키워드는 워드클라우드 형태로 시각화되며, 사용자는 한눈에 어떤 키워드가 많이 언급되었는지 확인할 수 있다.

키워드는 글자 크기 차이를 통해 빈도를 직관적으로 구분할 수 있도록 구성되었다.

예를 들어 많이 등장한 용어는 크게 표시되고, 적게 등장한 용어는 작게 표시되어 시각적으로 비교가 가능하다.

또한 사용자가 키워드를 클릭하면, 해당 키워드가 포함된 뉴스 기사 목록으로 즉시 이동할 수 있다.

이를 통해 사용자는 별도의 검색 과정 없이 특정 키워드와 관련된 뉴스만 빠르게 모아 볼 수 있으며, 관심 있는 이슈를 효율적으로 탐색할 수 있다.

이 기능은 단순한 뉴스 나열이 아니라, 하루 동안 어떤 경제 흐름이 있었는지를 요약해 보여주는 역할을 하며 사용자의 정보 탐색 시간을 줄여준다.



## 7. 단답형 퀴즈

단답형 단어 퀴즈

Q1. 해설: (1) AI는 인간의 학습 능력과 추론 능력, 지각 능력, 자연언어의 이해 능력 등 인간의 지적 능력을 컴퓨터 프로그램으로 구현하는 과학기술이다 (2) 1956년 다트머스대학에서 열린 회의에서 존 매카시가 '□□□□(Artificial Intelligence)'이라는 용어를 처음 사용했으며, □□□□을 '지능적인 기계를 만드는 과학 및 공학(The science and engineering of making intelligent machines)'으로 정의했다 (3) 지금은 컴퓨터와 사람이 직접 바둑이나 장기를 둔다든지, 수리적인 정리를 증명하는 프로그램 등이 개발되어 있으며, 의학 분야에서는 의사의 진단을 돕기 위한 진단 시스템이 개발되어 실제로 환자 진료에 이용되고 있다

인공지능

조성 힌트: ㅇ ㄱ ㄴ ㄷ ㄹ

Q2. 해설: (1) 이종기록을 단일기록으로 결합하여 둘 또는 그 이상의 목록이나 파일을 결합한다 (2) □□은 일상적으로 바람직하지 않은 이름을 제거하는 제거과정에 관련하여 행해진다 (3) 재무적으로 두 회사가 결합하여 단지 한 회사만이 법적 실체로 존속한다.

합병

조성 힌트: ㅇ ㅂ

정답 보기

Q3. 해설: (1) □□은 □□회사 에 출자를 한 주주가 주주로서 회사에 대해 가지는 법률상의 권리·의무 단위인 주주권 (株主權)으로서의 뜻이 있다 (2) □□의 본질에 대하여는 주주가 주주라는 지위에서 가지고 있는 권리와 의무가 합쳐져 단일한 권리로 되었다는 설, 그리고 주주가 주주의 자격에서 가지고 있는 많은 권리의무가 집합하였다는 설, 또 주주의 권리의무가 생기게 되는 법률상의 지위 또는 자격이라는 설 등이 있으나 주주의 회사에 대한 법률상의 지위라고 해석하는 것이 가장 타당할 것이다 (3) 주주는 이러한 지위에 근거해서 회사에 대해서 여러가지 권리를 가지게 되는데, 주주의 권리는 주주가 소유하는 □□의 수에 비례하여 정해진다

자산

조성 힌트: ㅇ ㅅ ㅈ ㅊ ㅋ ㆁ

정답 보기

나의 레벨 현황

현재 등급: 골드

브론즈

다음 단계: 동적티널

목표: 2000점 (현재: 0점)

실버

다음 단계: 동적티널

목표: 2000점 (현재: 0점)

골드

다음 단계: 동적티널

목표: 2000점 (현재: 0점)

다음 단계(동화): 동적티널

목표: 2000점 (현재: 0점)

본 시스템에서는 사용자가 저장해둔 단어를 바탕으로 학습할 수 있도록 단답형 퀴즈 기능을 제공한다. 퀴즈 문제는 사용자가 단어장에 추가한 단어 중 하나를 선택하여, 해당 단어의 해설(설명)을 제시하는 방식으로 출제되며, 사용자는 해설을 보고 단어를 직접 입력하여 정답을 맞혀야 한다.

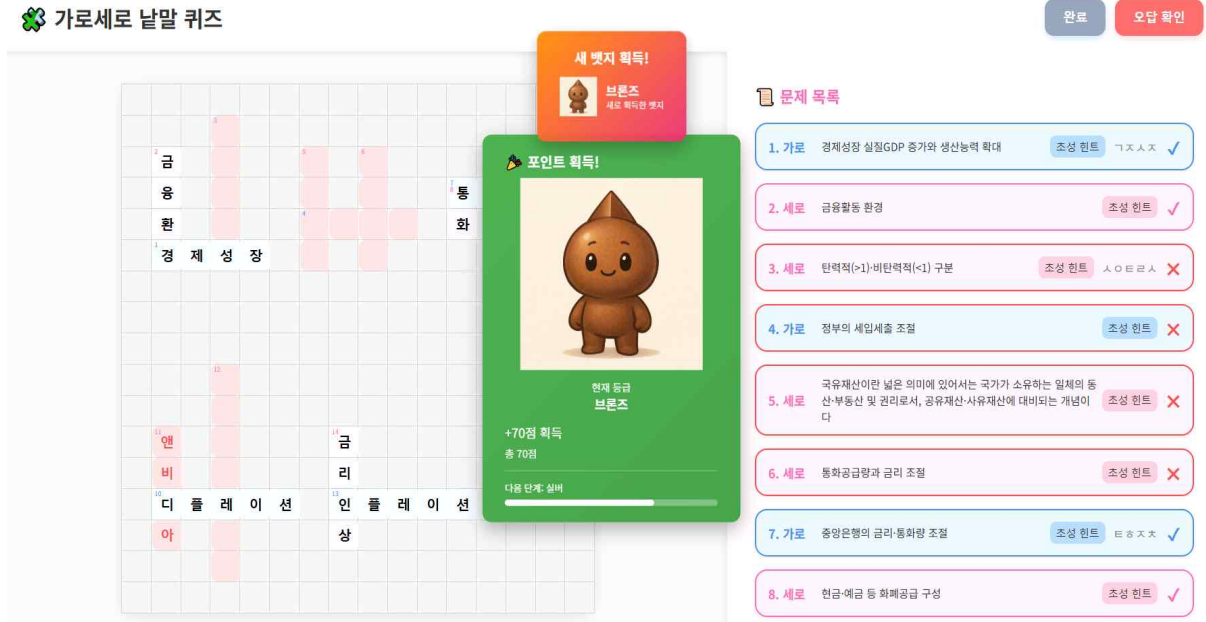
퀴즈 진행 과정은 다음과 같다:

1. 단어장에 저장된 단어 중 하나가 선택되고, 해당 단어의 설명(뜻)이 문제로 표시된다.
2. 사용자는 제공된 해설을 참고하여 정답 단어를 입력한다.
3. 정답이 떠오르지 않을 경우, 초성 힌트 버튼을 클릭하여 도움을 받을 수 있다.
4. 정답을 맞히면 해당 단어는 단어장에서 자동으로 제거되며, 사용자의 학습 진행 상황이 반영된다.
5. 정답을 맞출 때마다 사용자 캐릭터의 경험치가 증가하여, 학습의 성취감을 느낄 수 있도록 하였다.

단답형 퀴즈는 단순히 단어를 암기하는 방식이 아니라,

'단어 해설을 보고 → 단어를 입력하는' 능동형 학습 구조를 가지고 있어 사용자가 경제·금융 용어를 직접 기억하고 사용할 수 있도록 도움을 준다.

## 8. 가로세로 낱말 퀴즈



본 시스템에서는 사용자가 저장해둔 단어들을 바탕으로 학습할 수 있도록 **가로세로 낱말 퀴즈** 기능을 제공한다.

이 기능은 단어장 내부에 저장된 경제·금융 용어를 조합하여 퍼즐 형태의 문제를 생성하며, 각 단어의 해설(뜻)이 문제의 힌트로 제시된다. 사용자는 해설을 보고 정답 단어를 가로 혹은 세로 방향으로 입력하여 퍼즐을 완성해 나간다. 퀴즈 진행 방식은 다음과 같다:

1. 단어장에 저장된 단어들이 자동으로 선택되어 퍼즐 형태로 배치된다.
2. 우측에는 각 단어에 대한 문제 목록과 힌트(해설)이 제공된다.
3. 사용자는 힌트를 참고하여 가로 혹은 세로 방향에 해당하는 칸에 정답을 입력한다.
4. 정답을 맞히면 해당 단어는 정답 처리되며, 퀴즈 진행 상황이 즉시 반영된다.
5. 여러 문제를 풀어 퍼즐을 완성하면 포인트를 획득할 수 있으며, 포인트는 캐릭터 경험치로 반영되어 레벨이 올라간다.

가로세로 낱말 퀴즈는 단순 암기 방식이 아니라,

“단어 해설을 보고 → 생각하고 → 직접 입력하는 방식”

으로 구성되어 있기 때문에, 사용자가 단어를 자연스럽게 기억할 수 있도록 돕고 학습의 재미를 제공한다.

## 9. 마이페이지

### 마이페이지

프로필

#### 플래티넘을 향해

아이디: platinum\_seeker

시스님, 오늘도 금융 한 걸음

보유 포인트
750 P

50%

이번 주 학습률  
목표: 10개 뉴스

가장 많이 본 키워드:

#미국
#보안
#중국

내가 본 뉴스

최근 본 순

**美, MS에 UAE AI집 수출 허용...엔비디아 집, 중동행**

미국 경제 수출 중국

한국경제 2025.11.03

**주가 급등하던 SK하이닉스 날벼락...투자주의종목 지정돼**

중국 지정 투자 하이닉스 장치

한국경제 2025.11.03

**MS, UAE AI 인프라에 11.3조 투자...엔비디아 블랙홀도 선적 승인받아**

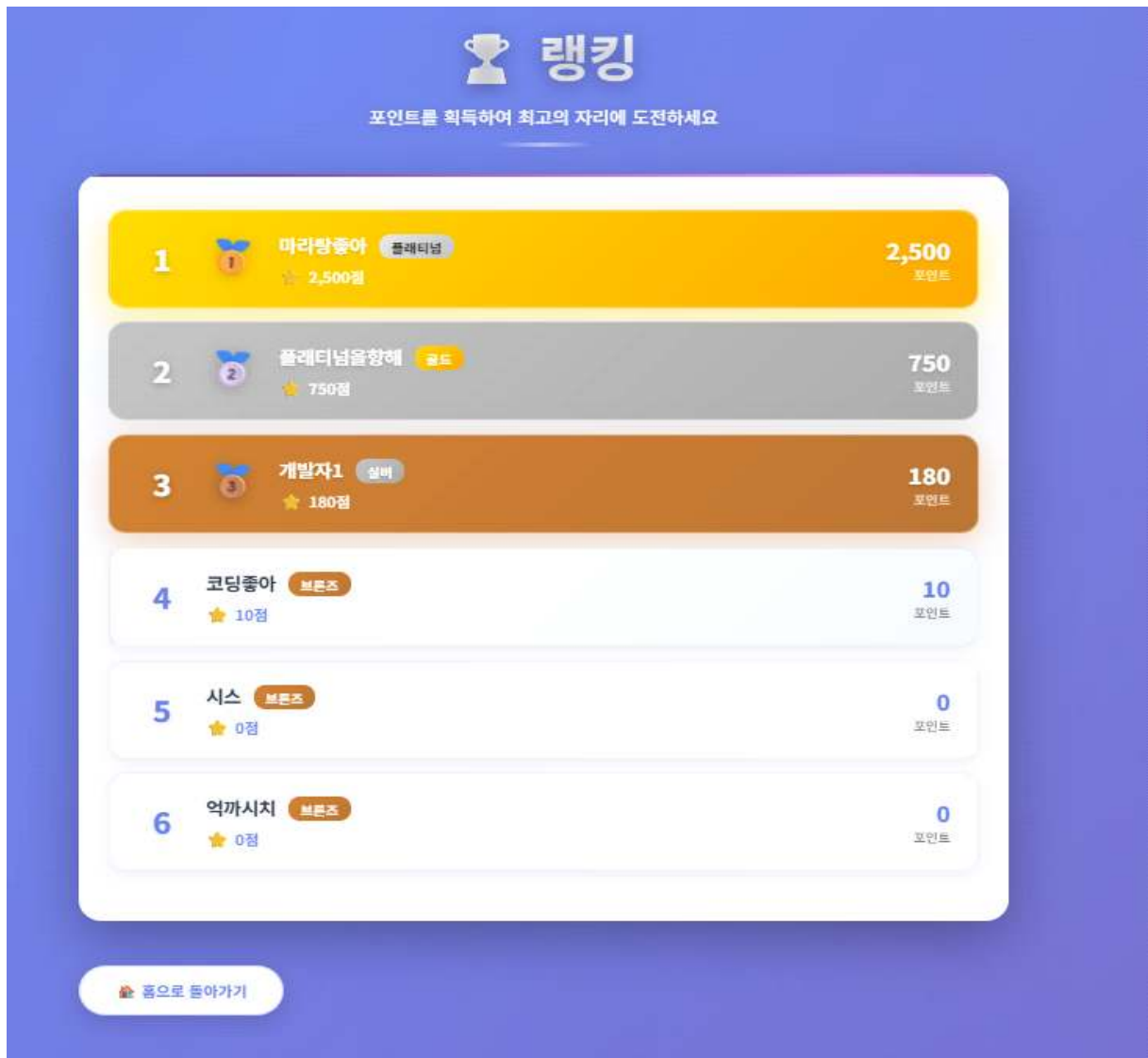
마이크로소프트 투자 보안 데이터 미국

한국경제 2025.11.03

< 이전
1 / 2 페이지 (5개)
다음 >

마이페이지는 사용자의 활동 현황과 학습 진행 상황을 한눈에 확인할 수 있는 공간이다. 이 화면에서는 사용자가 최근에 열람한 뉴스 목록과 해당 뉴스에서 추출된 주요 키워드가 함께 제공되어, 사용자가 어떤 분야에 관심을 가지고 있는지 쉽게 파악할 수 있다. 또한 사용자 캐릭터의 티어 등급, 현재 경험치, 누적 포인트가 함께 표시되며, 이를 통해 사용자가 얼마나 학습에 참여하고 있는지 시각적으로 확인할 수 있다. 포인트와 경험치는 뉴스 열람, 단어 학습, 퀴즈 활동을 통해 누적되며, 일정 수준에 도달하면 캐릭터가 성장하거나 티어가 상승하는 구조로 되어 있다. 마이페이지는 단순한 정보 제공을 넘어, 사용자가 스스로의 학습 성과를 인식하고 계속해서 참여할 수 있도록 동기를 부여하는 핵심 기능이다.

## 10. 게임 랭킹 시스템



랭킹 페이지는 사용자들의 학습 참여도를 높이기 위한 경쟁형 시스템으로, 사용자가 활동을 통해 획득한 포인트를 기준으로 전체 순위가 산정된다.

포인트는 뉴스 학습, 퀴즈 정답, 단어장 활동 등 다양한 학습 참여를 통해 얻을 수 있으며, 누적된 포인트는 실시간으로 업데이트된다.

상위 랭킹에 오른 사용자는 자신의 닉네임과 함께 현재 보유 포인트, 그리고 사용자의 캐릭터 티어가 표시된다. 또한, 일정 기준의 포인트를 달성할 경우 배지를 자동으로 획득하게 되며, 이를 통해 사용자는 자신의 성취를 시각적으로 확인할 수 있다.

배지는 브론즈, 실버, 골드 등 단계별로 나뉘어 있어 자연스럽게 더 높은 랭킹과 다음 단계의 배지를 목표하게 된다.

랭킹 시스템을 통해 경쟁 요소가 추가되어 사용자들은 학습에 대한 동기와 흥미를 유지할 수 있으며, 다른 사용자와의 성과 비교를 통해 지속적인 참여를 유도한다.

## 라. 개선 방안

현재 시스템은 단답형과 가로세로 퀴즈 중심으로 구성되어 있어 경제·금융 용어 학습에 특화된 프로그램으로 보기 어렵다. 이러한 한계를 개선하기 위해서는 뉴스와 경제 데이터를 유기적으로 연계한 고도화된 학습 기능이 필요하다. 우선, 페이지 내에 GDP, 물가상승률, 환율, 금리 등 주요 경제지표를 실시간으로 확인할 수 있는 기능을 추가하고, 사용자가 뉴스를 열람할 때 기사 내에 해당 지표가 언급되면 자동으로 최신 수치를 함께 보여주는 방식으로 정보를 직관적으로 제공할 수 있다. 이를 통해 사용자는 별도의 검색 과정 없이 기사 내용을 빠르게 이해할 수 있으며, 학습 효율성이 높아진다. 또한 기존의 단순 암기형 퀴즈를 넘어 뉴스 문맥에 따라 자동으로 생성되는 문항이나 간단한 해설형 문제를 제공함으로써 실질적인 이해 중심 학습이 가능해진다. 사용자가 정답을 맞히거나 해설을 열람할 때마다 관련 용어가 자동으로 복습 리스트에 추가되어 반복 학습이 이루어지도록 하면 학습 지속성이 강화된다. 더불어 사용자별 학습 기록과 관심 분야를 기반으로 개인 맞춤형 콘텐츠를 추천하고, 오답률이나 참여 빈도에 따라 난이도를 조절하는 기능을 추가하면 학습 몰입도를 더욱 높일 수 있다.

한편, 현재 프로그램은 경제 분야에 국한되어 있어 정치, IT, 국제 등 다양한 뉴스 분야를 포괄적으로 다루기에는 한계가 존재한다. 이에 따라 각 분야별 핵심 용어와 주요 지표를 포함한 데이터 베이스를 단계적으로 구축한다면, 뉴스의 주제와 맥락에 따라 관련 해설과 배경지식을 자동으로 제공하는 종합형 학습 플랫폼으로 확장할 수 있다. 이러한 다분야 연계형 구조는 사용자가 특정 기사나 용어를 접할 때, 해당 분야의 기초 개념부터 심화 정보까지 연속적으로 학습할 수 있도록 지원해 학습 효율성을 높인다.

또한 학습 참여율을 높이기 위해 도입된 랭킹 시스템은 단순한 순위 표시 기능을 넘어, 사용자의 학습 활동(예: 퀴즈 참여, 용어 열람, 해설 확인 등)에 따라 점수를 부여하고 일정 수준을 달성하면 배지나 추가 학습 콘텐츠, 개인화된 추천 자료 등을 제공하는 보상 기반 시스템으로 발전시킬 수 있다. 이를 통해 사용자는 자신의 학습 성과를 시각적으로 확인하며 지속적인 동기를 부여받게 되고, 경쟁과 성취를 기반으로 한 자발적 참여를 유도할 수 있다. 결과적으로 이러한 구조적 개선은 본 시스템을 단순한 용어 퀴즈 프로그램이 아닌, 경제를 포함한 다양한 시사 영역의 지식을 통합적으로 학습하고 이해할 수 있는 지식 확장형 학습 플랫폼으로 발전시키는 기반이 될 것이다.

### 3. 과제의 향후 계획

#### 가. 활용 방안

본 시스템은 경제 뉴스를 중심으로 주요 용어와 지표를 자동 인식하고, 이에 대한 해설을 제공하는 학습 지원 플랫폼으로 활용될 수 있다. 사용자는 뉴스 기사를 읽는 과정에서 어려운 경제·금융 용어나 개념을 별도의 검색 과정 없이 즉시 이해할 수 있으며, 이를 기반으로 학습 데이터가 자동 축적된다. 축적된 학습 이력은 개인별 수준에 맞춘 맞춤형 퀴즈나 복습 콘텐츠 제공에 활용되어, 반복 학습을 통한 개념 정착을 돕는다. 또한 주요 경제지표를 실시간으로 연동해 보여줌으로써 뉴스의 맥락과 실제 데이터 간의 연관성을 직관적으로 파악할 수 있다. 이외에도 정치, IT, 국제 등 다양한 분야의 뉴스 데이터베이스를 확장 구축하면, 경제 외의 다양한 시사 이슈에 대해서도 동일한 학습 환경을 제공할 수 있다. 학습 과정에서 획득한 점수는 랭킹 시스템과 연계되어 사용자 간 경쟁을 유도하며, 상위권 사용자에게는 추가 콘텐츠나 배지 제공 등 보상 체계를 운영함으로써 학습 동기를 강화할 수 있다. 이를 통해 본 시스템은 단순한 뉴스 열람 도구가 아니라, 실시간 시사 학습이 가능한 지식 기반 학습 플랫폼으로 활용될 수 있다.

#### 나. 기대 효과

본 시스템을 활용하면 사용자는 복잡한 경제·시사 뉴스를 보다 쉽게 이해하고, 자연스럽게 금융 및 사회 전반에 대한 문해력을 향상시킬 수 있다. 자동 해설 기능과 연계된 학습 구조를 통해 비전공자도 전문 용어나 지표의 의미를 쉽게 습득할 수 있으며, 반복 노출을 통해 실질적인 학습 효과를 기대할 수 있다. 또한 개인별 학습 데이터를 기반으로 맞춤형 콘텐츠를 제공하므로 효율적인 학습 환경이 조성된다. 랭킹 시스템과 보상 체계는 학습 참여율을 높여 지속적인 이용을 유도하며, 사용자 간의 경쟁을 통해 학습 커뮤니티 활성화도 가능하다. 나아가 다양한 분야의 뉴스 데이터를 통합함으로써 경제뿐만 아니라 정치, IT, 국제 등 폭넓은 영역의 시사 이해도를 높여, 종합적 사고력과 정보 분석 능력을 향상시키는 효과를 거둘 수 있다. 결과적으로 본 시스템은 뉴스 소비를 단순 정보 습득을 넘어 능동적 학습 활동으로 전환시키는 교육적 플랫폼으로 자리매김할 것이다.

#### 4. 참고문헌

#### 5. 별첨