

슬라이드 1. 표지

안녕하십니까. 클라우드컴퓨팅전공 박정현입니다.
제가 발표할 프로젝트는 IT/Science News Hub로, 여러 매체에 흩어진 IT/과학 기사를 한 곳에서 모아 요약과 검색을 제공하는 웹 애플리케이션입니다.
2주 동안 진행한 MVP 결과를 발표하겠습니다.

슬라이드 2. 프로젝트 개요 & 문제 정의

현재 대한민국의 언론 매체는 한국언론진흥재단에 속해 있는 참여 언론사만을 기준으로 봤을 때 127개에 달합니다. 이처럼 IT/과학 기사가 여러 매체에 분산되어 있고, 매체별 편향성과 부족한 요약 기능 때문에 독립적이고 효율적인 정보 전달이 어렵습니다.

이러한 문제를 해결하기 위해, 저는 여러 매체의 IT/과학 기사를 자동으로 수집하고, 통합된 플랫폼에서 검색과 간단한 요약 기능을 제공하는 뉴스 허브 서비스를 구축하고자 했습니다.

슬라이드 3. 핵심 기능과 해결 방안

이 프로젝트의 핵심 기능은 세 가지입니다.

첫째, **자동 기사 수집**입니다.

다양한 매체의 IT/과학 기사를 GitHub Actions로 주기적으로 자동 수집합니다.

둘째, **통합 검색 시스템**입니다.

PostgreSQL Full Text Search와 pg_trgm 인덱스를 활용해 원하는 기사를 빠르고 정확하게 검색할 수 있습니다.

셋째, **스마트 요약 제공**입니다.

기사 카드를 썸네일과 함께 간단한 규칙 기반 요약으로 제공하고, 추후에는 LLM 요약으로 확장할 계획입니다.

슬라이드 4. 시스템 아키텍처

시스템 아키텍처를 설명드리겠습니다.

이 프로젝트는 네 가지 레이어로 구성되어 있습니다.

먼저 왼쪽의 **Automation 영역**에는 Python 스크립트가 GitHub Actions에서 동작합니다. 여기서 여러 매체의 기사를 주기적으로 스크래핑하고 Supabase에 자동으로 적재합니다.

Database 영역은 Supabase를 사용합니다. 일간 기사와 주간 기사를 저장하고, 이 데이터는 unified articles view에 통합되어 검색에 활용됩니다.

Server 영역은 Express API를 사용합니다. API는 크게 두 가지 역할을 합니다. 하나는 daily, weekly, sidebar 요청에 따라 기사를 가져오는 기사 조회 API, 다른 하나는 unified articles view를 활용하는 검색 API입니다.

마지막으로 **Frontend 영역**입니다. next.js 기반으로 구현된 UI는 Daily/Weekly 카드 기사 UI와 검색 결과 화면을 제공합니다.

정리하면, 사용자가 브라우저에서 daily나 weekly를 탭하거나 검색을 수행하면 frontend가 API 서버로 요청을 보내게 되고 API 서버는 Supabase에 필요한 데이터를 조회해 JSON 형태로 반환하고, 이 결과가 기사 카드 UI나 검색 결과 화면에 표시됩니다.

슬라이드 5. 핵심 기술적 특징

핵심 기술은 세 부분입니다.

첫째, **검색 기능**입니다.

PostgreSQL FTS를 활용해 빠르고 정확한 기사 검색을 구현했습니다.

둘째, **마이그레이션 관리**입니다.

Supabase를 활용해 데이터베이스 스키마를 체계적으로 관리했습니다.

셋째, **자동화 시스템**입니다.

GitHub Actions 기반 배치 작업으로 기사 수집을 자동화했습니다.

슬라이드 6. 시스템 데모

지금부터 MVP 데모를 보여드리겠습니다.

먼저 Daily 탭입니다.

하루 단위로 수집된 IT/과학 기사를 카드 형태로 보여주며, 썸네일, 제목, 요약이 함께 표시되어 핵심 내용을 빠르게 확인할 수 있습니다.

다음은 Weekly 탭입니다.

주간 단위 기사를 모아 한 주의 흐름과 주요 트렌드를 쉽게 파악할 수 있습니다.

오른쪽 사이드바를 통해서 원하는 날짜를 선택해 해당 시점의 기사를 바로 파악할 수 있습니다.

마지막으로 검색 기능입니다.

예를 들어 '반도체'를 검색하면 PostgreSQL FTS 기반으로 Daily와 Weekly 기사에서 관련 기사를 신속하게 찾아 카드 형태로 제공합니다.

보신 것처럼, 이번 MVP는 daily/weekly 기사 제공, 날짜별 탐색, 키워드 검색까지 기본 기능을 모두 구현했습니다.

다음으로는, 현재 한계와 개선 필요 사항을 설명드리겠습니다.

슬라이드 7. 현재 한계와 개선 필요 사항

2주 동안 진행한 프로젝트의 MVP라 몇 가지 한계가 있습니다.

첫째, 주제별 분류 시스템입니다.

여러 매체에서 들어온 기사를 자동으로 분류하고, 동일한 주제끼리 묶을 수 있는 기능이 구현되지 않았습니다.

둘째, 고도화된 요약 기능입니다.

현재는 규칙 기반으로 간단한 요약만 제공되며, LLM을 활용한 요약은 적용되지 않았습니다.

이 부분들은 앞으로 개선해야 할 과제들입니다.

슬라이드 8. 차별성과 기대효과

아직 미완성된 부분이 있지만 현재 구현된 기능들을 기반으로 확장한다면, 차별성과 기대효과를 기대할 수 있습니다.

첫째, **전문 분야 특화**입니다.

IT와 과학 기사에만 특화된 허브로서, 이 분야에서 전문성을 확보할 수 있습니다.

둘째, **사용자 경험**입니다.

하나의 플랫폼에서 여러 매체의 기사를 모아 효율적으로 확인할 수 있습니다.

셋째, **트렌드 분석**입니다.

빠르게 변화하는 기술 트렌드와 시장 동향을 신속하게 파악할 수 있습니다.

마지막으로, **무한 확장성**입니다.

PostgreSQL 기반으로 향후에는 pgvector 기반 벡터 검색이나 LLM 요약 같은 차세대 기술을 적용해 더 발전된 뉴스 플랫폼으로 확장할 수 있습니다.

슬라이드 9. 향후 발전 계획

이번 프로젝트는 MVP 단계에서 기본적인 기사 수집, 검색, 요약 기능을 구현하는 데 집중했습니다. 앞으로의

발전 방향은 두 가지입니다.

첫 번째 **주제별 클러스터링**을 도입해 여러 매체에서 들어온 기사를 자동으로 분류하고, 동일한 주제끼리 묶을 수 있도록 하는 것입니다.

두 번째 **통합 요약**을 적용합니다.

묶인 기사를 맥락적으로 분석해 중복된 내용을 제거하고, 하나의 고품질 요약 기사로 정리해 사용자에게 제공합니다.

슬라이드 10. 마무리

감사합니다.