



Different News

▼ 기획안

[Different News] 임시명

대한민국의 큰 이슈는 대부분 정치적 이슈와 밀접하게 맞닿아 있다.

정치에서 나뉘는 진보(좌파, 사회 전반의 변화나 발전을 추구하는 것을 의미), 보수(우파, 기존의 사회 질서나 전통적인 가치를 유지하는 것을 의미)에 따라 정치에 큰 영향을 끼치는 대한민국 언론사의 성향 또한 구분되어 있다.

한 키워드에 대해(한 이슈에 대해) 자신이 추구하는 방향의 정당과 비슷한 신념의 언론사의 뉴스를 찾아보는 것도 좋지만, 같은 키워드에 대한 이념적 생각을 가진 자들의 뜻도 무엇인지 알아보는 **것도 중요하다.**

다른 색의 사람들이 한 공간에 모여 서로의 생각(뉴스)을 읽어보고, 각 생각(뉴스)에 대한 서로의 생각 또한 알 수 있는 웹 사이트 구축이 목표이다.

즉, 목적은 갈라치기가 아닌 같은 생각을 가진 사람의 소리를 듣는 동시에 다른 생각을 가진 사람의 소리 또한 들어보아 **정보의 불균형을 해소하는 것이다.**

초안

node.js + python + db(MySQL)

Different News

설명

- 좌(경향, 한겨레, 오마이뉴스), 우(조선 일보, 중앙 일보, 동아 일보)의 각 RSS를 활용하여 인기 토픽에 대한 양 측 기사들을 리스트업 해주고, 해당 토픽에 대한 생각을 댓글창(라이트), 토론방(헤비)을 통해 공유.

제품 핵심

- AI를 활용한 데이터 분석으로 핫 이슈 5~10개 추출 → 관리자가 최종 토픽 선별
- 선별된 토픽에 대한 좌/우 기사 분류

성장

- 입소문→임계치 도달 시 제휴/라이선스 협상.

=====

▼ IA(Information Architecture)

1. 홈

- 상단: 사이트이름/간단한 홈페이지 소개/로그인, 마이페이지
- 본문: “인기 토픽 top n” 카드 그리드(토픽명, 간단 설명/핵심 키워드 3개, 생성 시각, hot, 썸네일 등 설명) 무한스크롤
- 푸터: 약관, 운영원칙, 신고/문의 등

2. 각 토픽 페이지

- 왼쪽 사이드: 인기 토픽 n개 리스트업(볼륨, 출처 다양성, 좌/우 균형, 조회수 등으로 가중치 부여 정렬)
- 오른쪽 사이드: 참고 언론사 리스트업 or 광고 (더 생각)
- 헤더: 토픽명, 간단한 설명(중도의 입장), 생성 시각

▼ 예시



- 본문: 좌/우 2열 레이아웃(PC), 모바일은 탭 전환(좌/우)
 - 각 열: 언론사 구분 섹션→기사 카드 리스트(원제목, 게재일-언론사명, 바로가기 링크)
 - 각 언론사 별or최신순 페이지네이션 또는 무한스크롤
 - 하단: 참여 영역의 댓글(라이트)
 - 짧은 의견, 추천/비추천 기능, 정렬(최신/추천순), 신고
 - 추가 기능⇒성향 커밍아웃 라벨, 클린 뱃찌 등
3. 토론방 시작 페이지(토론방 시작 페이지로 넘어갈 수 있는 trigger를 만들어야 함.)
- 왼쪽 사이드: 인기 토픽 10개 리스트업
 - 오른쪽 사이드: 참고 언론사 리스트업 or 광고 (더 생각)

- 헤더: 간단한 설명(토론 및 토론방 설명)
- 본문: 토론방 생성
 - 토론 주제(선택 or 생성), 시작 시각, 참여자 인원 제한
 - 리스트업 된 hot 토픽에 대한 예시 토론 주제 3~5개 중 선택 가능

4. 토론방(아직 미흡+다른 의견)

- 오른쪽 사이드: 참여자 이름 리스트업
- 본문: 채팅
- 하단: 관전자(챗 x. but 엄지 등 이모지?로 표현)
 - 규칙 설명란(ex. 인사로 시작 끝, 비방 및 욕설 금지 등)

▼ 토론의 주요 원칙

- **추정의 원칙:** 특정 주장이 증명되지 않은 한, 그 주장이 참이라고 단정해서는 안 됩니다.
- **평등의 원칙:** 모든 참여자에게 동등한 발언 기회가 주어져야 합니다.
- **상호 존중의 원칙:** 상대방을 존중하고, 강압이나 협박보다는 설득으로 문제를 해결해야 합니다.
- **결과 승복의 원칙:** 토론 결과에 대해 승복하고, 토론을 통해 성찰하는 자세가 필요합니다.

• 토론 룰 ex)

- 토론 시작 시 토론자들은 모두 “매너있는 토론을 진행하겠습니다” 치고 시작하게 하기
- 자신의 주장 띄워놓기(옆 자신의 이름 아래)

5. Q&A / 메일: ~기사 다뤄주세요, 채팅창 점검 해주세요 등

5. 게시물(마이너 커뮤니티가 될 수도 있음)

▼ 한 기사 예시 구조

- 출처: 파비콘+언론사 명(파비콘 URL 캐싱 필요)
- 제목: 기사의 원제목(원문 바로가기: 직접링크)
- 요약(1~3줄): 사실 요약 + 짧막한 코멘트(왜 읽을 가치가 있는지) (저작권)
- 키워드: 3~5개 (저작권)
- 게재일(ex 23분 전, 1시간 전, 어제 / **9월 16일 17:00**)

▼ 예시(기사)

대통령실

대한민국
대통령실 “우리 기업 손해 보는 합의안 서명 불가”…대미 관세협상 관련 거듭 강조

18시간 전

관련 콘텐츠

조선일보
최임 두 달, 러트릭만 20번 만난 김정관… “책상 치고 목소리 올라가기 도”
23분 전

v.daum.net
통상본부장, 미 무역대표와 회담…‘대미투자 이견’ 좁히기 주력
3시간 전

경향신문
산업장관 “3500억불 대신 관세 보조금 주면 어떨까 생각 많이해”
15분 전

+관련 콘텐츠 누르면 각 성향 언론사들 홈페이지 리스트업(모달?)

+시각화가 중요하다고 생각. 썸네일 필요(뉴스토어에서 사진만 구매 가능(보도 사진 섹션)-추후 논의)

▼ AI 썸네일 가이드라인

- 안전 수칙(핵심)
 - **사람 얼굴, 실존 인물 금지:** 정치인, 연예인, 기자 등 식별 가능한 인물을 닮게 그리거나 합성하면 초상권/명예훼손 리스크가 커짐. 가능하면 아이콘, 추상 일러스트로 대체.
 - **로고, 브랜드, 언론사 마크 금지:** 썸네일에 JTBC, KBS, 삼성 로고 같은 제 3자 상표/마크를 넣지 않기. (혼동, 상표 리스크)
 - **사실 오인 유발 금지:** 범죄/사고 기사에 특정 회사/기관을 연상시키는 건물, 유니폼, 색채/패턴을 쓰지 않기. 묘사는 중립적, 추상적으로.
 - **원본 자동 프리뷰 차단:** 링크의 OG 이미지/문구가 자동으로 뜨지 않게 하고, og:image 는 내가 만든 썸네일로 명시.
 - **표시 문구:** 페이지 어딘가에 작게 “썸네일은 AI 생성 일러스트입니다.” 정도의 고지(특히 논쟁 이슈일수록 권장됨.)
 - **저작권, 약관 체크:** 사용하는 이미지 생성 모델의 이용약관에서 “출력물의 사용권”을 확인
 - **이의제기 대응:** 문의/클레임 시 신속히 교체, 삭제하는 절차를 준비.
- 실무 팁
 - **프롬프트 가드레일:**
“Minimalist illustration of [주제], no faces, no people, no logos, no brands, no text, abstract icons, neutral colors, 1200×630”
 - **일관된 디자인 시스템:** 카테고리별 컬러/아이콘 세트(경제=파랑, 정치=보라...)를 정해 혼동 없는 ‘우리만의’ 시각 언어 만들기.

- 접근성/메타: 1200×630(또는 1200×628) 권장, alt 텍스트 "(일러스트) [주제] 관련 추상 아이콘" 같이 명확히.
- 로그 보관: 생성 프롬프트·시드·버전을 저장(사후 분쟁 시 설명 가능).

▼ 참고 사이트

- 오마이뉴스/논쟁
 - https://www.ohmynews.com/NWS_Web/Debate/index.aspx
 -
- 한국경제/생글생글
 - <https://sgsg.hankyung.com/series/1509001006>
 - 활성화 X

토픽 선정 파이프라인(중요)

단계 & 규칙

1. 수집(크롤러/RSS/검색트렌드)

- RSS 폴링(15분 간격) → articles UPSERT(키: domain + normalized_title + published_at)
- 시간 표준: KST, 상대·절대 시각 모두 저장(rel , abs)

2. 임베딩(제목) → 군집화 (AI 허용 소스만)

- article.ai_allowed=TRUE (AI 허용)인 것만 제목 임베딩 생성 → 최근 24~48h만 대상으로 HDBSCAN/BERTopic
- 가드레일
 - 최소 클러스터 크기: ≥3개 기사
 - 유사도 평균(클러스터 내) ≥0.80 미만이면 보류 큐
 - 동일 도메인 비중 >60%면 "과도 단일 출처" 경고

1. 군집명 자동 생성(일반 주제명)

- 출처 문장 복붙 없이 일반 주제명 생성(예: "정년연장 논의", "민생쿠폰 지급")
- 후보 3개 자동 제안 → 운영자 선택 (선택 전에도 임시명으로 표시 가능)

2. 좌/우 대표 기사 최소 3건 보장

- 각 토픽에 대해 좌/우(매체 성향 테이블로 결정) 각 ≥ 3 건 충족 시 자동 “준비완료”
- **미충족 시 보강 절차**
 - AI 허용 기사 풀에서 동일 키워드/엔티티로 재검색(최근 72h까지 확장)
 - 그래도 부족하면 운영자에게 알림: No-AI 기사 수동 연결 요청(좌/우 불균형 배지)

3. 중복·낚시성·오보 리스트 필터

- **블랙리스트 도메인** 차단(오보·낚시성 기록된 매체)
- 제목 패턴(“충격/헉/ㄷㄷ/클릭”) 차감 또는 제외
- **동일 URL/제목 중복 제거**(UTM 제거, canonical 정규화)

4. 운영자 최종 승인

- 토픽 상태: `draft → review → published`
- 화면에서 확인: 토픽명 / 좌·우 기사 수 / 불균형 배지 / 블랙리스트 제거 내역
- [병합] 후보 / [분할] 미리보기 제공(자동 제안은 내부만 사용, 숫자 노출 X)

| 표현 원칙(사용자 화면): 원제목·출처·원문링크·게재시각 **만** 노출. (요약·재제목·키워드 없음)

아키텍처(YES-AI) – 구현 디테일

컴포넌트

- **Ingest(NestJS)**: RSS 폴링 \rightarrow `articles` UPSERT
- **Embed Worker**: `ai_allowed=TRUE` 만 제목 임베딩 \rightarrow `article_vectors`
- **Clustering Batch**: 15분 단위 \rightarrow `topic_candidates`, `topic_articles(AUTO)`
- **Scoring Batch**: 토픽 점수(불룸·출처다양성·좌우균형) \rightarrow `topic_scores`
- **Admin(React+Nest)**: 후보 검수/병합/분할/No-AI 수동 연결(**MANUAL**) \rightarrow Publish
- **Public API**: `/topics`, `/topics/:id` (좌/우 기사 목록 반환)

최종 FLOW

1. **수집**: RSS \rightarrow 제목/링크/매체/시간 저장(중복 제거)
2. **토픽 생성**: 허용 매체만 임베딩/규칙으로 묶어 토픽 만들기
 - 수집(RSS) \rightarrow 제목 임베딩(L2)
 - **DBSCAN 군집** \rightarrow 작은/품질 낮은 묶음 제거
 - 기존 토픽에 붙이거나 새 토픽 생성

- 좌/우 최소 보장 체크 → 부족 시 알림
3. 좌/우 라벨: 매체 성향 테이블로 기사에 좌/우 부여
 4. No-AI 기사 처리: Admin에서 드래그&드롭/선택으로 No-AI 기사를 특정 토픽에 직접 연결
 5. 검수/게시: 운영자가 병합·분할·게시 승인
 6. 사용자 화면: 이슈 상세에서 좌/우 칼럼에 기사 카드(파비콘+매체명 / 원제목(원문링크) / 게재 시각)

▼ 언론사 RSS 주소

- 경향 신문(제목 + 2~3 문장)

```
Feed("경향신문", "khan.co.kr", "LEFT", "https://www.khan.co.kr/rss/rssdata/politic_news.xml", "정치"),
    Feed("경향신문", "khan.co.kr", "LEFT", "https://www.khan.co.kr/rss/rssdata/economy_news.xml", "경제"),
    Feed("경향신문", "khan.co.kr", "LEFT", "https://www.khan.co.kr/rss/rssdata/society_news.xml", "사회"),
    Feed("경향신문", "khan.co.kr", "LEFT", "https://www.khan.co.kr/rss/rssdata/culture_news.xml", "문화"),
```

- 한겨레(AI 학습 및 활용 금지)

```
Feed("한겨레", "hani.co.kr", "LEFT", "https://www.hani.co.kr/rss/politics/", "정치"),
    Feed("한겨레", "hani.co.kr", "LEFT", "https://www.hani.co.kr/rss/economy/", "경제"),
    Feed("한겨레", "hani.co.kr", "LEFT", "https://www.hani.co.kr/rss/society/", "사회"),
    Feed("한겨레", "hani.co.kr", "LEFT", "https://www.hani.co.kr/rss/culture/", "문화"),
```

- 오마이뉴스

```
Feed("오마이뉴스", "ohmynews.com", "LEFT", "http://rss.ohmynews.com/rss/politics.xml", "정치"),
    Feed("오마이뉴스", "ohmynews.com", "LEFT", "http://rss.ohmynews.com/rss/economy.xml", "경제"),
    Feed("오마이뉴스", "ohmynews.com", "LEFT", "http://rss.ohmynews.com/rss/society.xml", "사회"),
```

```
Feed("오마이뉴스", "ohmynews.com", "LEFT", "http://rss.ohmynews.com/rss/culture.xml", "문화"),
```

- 조선일보

```
Feed("조선일보", "chosun.com", "RIGHT", "https://www.chosun.com/arc/outboundfeeds/rss/category/politics/?outputType=xml", "정치"),
```

```
Feed("조선일보", "chosun.com", "RIGHT", "https://www.chosun.com/arc/outboundfeeds/rss/category/economy/?outputType=xml", "경제"),
```

```
Feed("조선일보", "chosun.com", "RIGHT", "https://www.chosun.com/arc/outboundfeeds/rss/category/society/?outputType=xml", "사회"),
```

```
Feed("조선일보", "chosun.com", "RIGHT", "https://www.chosun.com/arc/outboundfeeds/rss/category/culture/?outputType=xml", "문화"),
```

- 중앙일보(우회 루트)

- 중앙일보는 자체 RSS서비스를 중단하여, Google News에서 발췌해서 RSS 긁어올 수 있음.

- 검색창에 ,로 키워드 입력 후 .com/뒤에 rss붙이기 ⇒ .com/rss/search?q\...

```
Feed("중앙일보", "joongang.co.kr", "RIGHT", "https://news.google.com/rss/search?q=site:joongang.co.kr%20정치&hl=ko&gl=KR&ceid=KR%3Ako", "정치"),
```

```
Feed("중앙일보", "joongang.co.kr", "RIGHT", "https://news.google.com/rss/search?q=site:joongang.co.kr%20경제&hl=ko&gl=KR&ceid=KR%3Ako", "경제"),
```

```
Feed("중앙일보", "joongang.co.kr", "RIGHT", "https://news.google.com/rss/search?q=site:joongang.co.kr%20사회&hl=ko&gl=KR&ceid=KR%3Ako", "사회"),
```

```
Feed("중앙일보", "joongang.co.kr", "RIGHT", "https://news.google.com/rss/search?q=site:joongang.co.kr%20문화&hl=ko&gl=KR&ceid=KR%3Ako", "문화"),
```

- 동아일보(AI학습 이용 금지)

```
Feed("동아일보", "donga.com", "RIGHT", "https://rss.donga.com/politics.xml", "정치"),
```

```
Feed("동아일보", "donga.com", "RIGHT", "https://rss.donga.com/economy.xml", "경제"),
```

```
Feed("동아일보", "donga.com", "RIGHT", "https://rss.donga.com/national.xml", "사회"),
```

```
Feed("동아일보", "donga.com", "RIGHT", "https://rss.donga.com/culture.xml", "문화"),
```

→Google News에서 인기 토픽에 대한 각 언론사의 RSS를 가져와서 AI돌려서 관리자(admin)에 보여준 후 관리자가 수동으로 선택한 뒤 사용자(api)에 리스트업.

+정치를 넘어 경제, 사회/국제까지 확장 가능.

▼ JSON 로직

- RSS URL 저장(예시)

```
{
  "sources": [
    {
      "name": "경향신문",
      "domain": "khan.co.kr",
      "side": "LEFT",
      "feeds": [
        "https://www.khan.co.kr/rss/rssdata/politic\_news.xml"
      ],
      "allow_ai_processing": true
    },
    {
      "name": "중앙일보",
      "domain": "joongang.co.kr",
      "side": "RIGHT",
      "feeds": ["<여기에 찾은 RSS URL>"],
      "allow_ai_processing": false
    }
  ]
}
```

- 표시용 JSON (사이드 카드 렌더링용-게시일 통합해야됨)

```
{
  "category": "검찰개혁",
  "updated_at": "2025-09-16T17:00:00+09:00",
  "items": [
    {
      "source": "경향신문",
      "source_domain": "khan.co.kr",
      "side": "LEFT",
      "title": "예시) 수사권·기소권 분리 논쟁 재점화...여야 공방 가열",
      ...
    }
  ]
}
```

```

    "url": "https://www.khan.co.kr/article/2025xxxxxxxx/",
    "published_at": "2025-09-16T10:30:00+09:00",
    "summary": null,
    "image_url": null
  }
]
}

```

- DB 삽입용 JSON (메타데이터 전용)

```

[
  {
    "category": "검찰개혁",
    "source": "경향신문",
    "source_domain": "khan.co.kr",
    "source_side": "LEFT",
    "origin_language": "ko",
    "section": "정치",
    "author": "홍길동",
    "content_type": "ARTICLE",           // ARTICLE | OPINION | EDITORIAL ...
    "url": "https://www.khan.co.kr/article/2025xxxxxxxx/",
    "url_canonical": "https://www.khan.co.kr/article/2025xxxxxxxx/",
    "published_at": "2025-09-16T10:30:00+09:00",
    "retrieved_at": "2025-09-16T10:45:12+09:00",
    "paywall": false,

    "ai_processing_level": "EMBED_SUMMARIZE", // NONE | EMBED | SUMMARIZE |
EMBED_SUMMARIZE
    "license_status": "UNKNOWN",           // OK | BLOCK | UNKNOWN
    "rss_guid": null,

    "title_text": "예시) 수사권·기소권 분리 논쟁 재점화...여야 공방 가열",
    "title_ai": null,                     // 재제목 사용 시
    "summary_ai": null,                  // 자체 요약 저장 시
    "keywords": ["검찰개혁", "수사권-기소권 분리", "검경수사권"],

    "og_image_url": null,                // 썸네일 수집 시
    "url_hash": "sha256:...",           // 선택
    "title_hash": "sha256:...",          // 선택
    "dedup_key": "sha256:khan.co.kr_수사권·기소권 분리 논쟁 재점화...여야 공방 가열
_2025-09-16"
  }
]
```

```
    }  
]
```

Node(API·Admin·크롤러) + Python(임베딩) 분리 마이크로서비스. 통신은 HTTP or 큐, 저장은
공용 DB/벡터DB

python 알고리즘/기술역할

1. 토픽 발견

K-Means Clustering

기사들을 주제별 그룹으로

묶는 역할

자연어 처리 (KoNLPy)

그룹의 이름(토픽명)을 짓는 역할

2. 기사 수집

코사인 유사도

선정된 토픽과 기사 간의 관련성을 측정하고 순위를 매기는 역할