

시맨틱 문장 분석을 활용한 뉴스 인터페이스 디자인

Designing a News Interface Using Semantic Sentence Analysis

이상아

SangAh Lee

서울대학교

지능정보융합학과

Department of Intelligence
and Information

Seoul National University
sangahlee718@snu.ac.kr

김정환

Junghwan Kim

서울대학교

지능정보융합학과

Department of Intelligence
and Information

Seoul National University
jhbale11@snu.ac.kr

송형우

Hyungwoo Song

서울대학교

지능정보융합학과

Department of Intelligence
and Information

Seoul National University
rotto95@snu.ac.kr

서봉원

Bongwon Suh

서울대학교

지능정보융합학과

Department of Intelligence
and Information

Seoul National University
bongwon@snu.ac.kr



그림 1. '뉴스 스냅샷' 개요: (i) 뉴스 기사 타임라인 뷰 (ii) 서브 토픽 뷰로 구성된다.

요약문

뉴스 산업의 발달 및 뉴스 소비 방법에 많은 변화가 생겼으나, 여전히 온라인에서 텍스트 기반의 뉴스 기사에 대한 소비는 꾸준히 이루어지고 있다. 하지만 뉴스 기사의 단편화 및 너무 많은 뉴스 기사의 수와 같은 문제점들은 디지털로 뉴스 기사를 읽는 독자들에게 불편함을 주고, 사건의 주요 내용과 이슈의 변화를 알아내기 어렵게 한다. 이러한 문제를 완화하고자 우리는 '뉴스 스냅샷' 시스템을 제안한다. 뉴스 스냅샷은 동일한 사건을 다루는 뉴스 기사들의 문장을 클러스터링하고 이를 바탕으로 뉴스 기사의 서브 토픽을 선정하여 제공한다. 지속되는 이슈에 대한 뉴스의 경우, 타임라인에 따라 주요 내용을 시각화하여 독자의 맥락 이해를 돕는다. 22 명의 참여자를 대상으로 수행한 사용자 평가를 통해 이 시스템의 효과 및 사용성을 검증하였다. 이와 같은 결과는, 뉴스 인터페이스 및 뉴스

관련 콘텐츠를 연구하는 후속 연구에서 유용하게 활용될 수 있을 것으로 보인다.

주제어

지능형 시스템, 온라인 뉴스 시스템, 뉴스 기사

1. 서론

디지털 미디어의 발전과 뉴스 산업의 변화로 뉴스를 제공하는 방법 및 독자들의 뉴스 소비 패턴에 다양한 변화가 나타나고 있다[1-3]. 하지만 여전히 텍스트 기반의 뉴스 소비가 가장 많이 이루어지고 있으며[2], 특히 한국에서는 포털 사이트와 같은 뉴스 집계 서비스를 이용한 뉴스 소비 비중이 가장 높다[1][3].

뉴스 기사가 온라인으로 제공되면서 뉴스의 발행 주기가 짧아지고 뉴스의 개수가 많아졌다. 수많은 뉴스 기사는 정보의 풍부함을 제공하는 반면, 일부 독자들은 어떤 기사를 선택하여 읽어야 할지에 대한 부담을 느끼게 되었다[4][5]. 또한 뉴스 보도의 실시간성 증대는 뉴스 기사의 길이를 단축시켰고, 이로 인한 정보의 단편화는 사건의 주요 내용과 그 흐름을 파악하기 어렵게 만든다[6][7].

기존 연구에서는 주로 시간이 지남에 따른 뉴스 범주의 분포 변화를 관찰했다[8][9]. 하지만 뉴스 기사의 이슈 관점에서 주요 내용이 어떻게 변해가는가에 대한 추적을 한 선행 연구는 드물다. 뉴스 기사 인터페이스 관련 선행 연구들은 대부분 필터버블이나 선택적 노출 관련하여 독자들의 행동 개선을 위한 목적으로 많이 수행되었다[10-12]. 뉴스 기사를 탐색하기 위한 인터페이스의 경우, 뉴스 기사 개별로 문장의 엔티티와 감정을 분석하는 것은 지원하지만[13][14], 대규모 뉴스 컬렉션에서 뉴스 기사의 이슈 별 토픽/서브 토픽 및 그 트렌드를 관찰하는 것을 지원하는 연구는 찾기 힘들다.

본 연구에서는 동일한 사건을 보도하는 많은 뉴스 기사에서 주요 내용을 추출하고 시간이 지남에 따른 이슈의 주요 흐름을 타임라인으로 제공하는 '뉴스 스냅샷' 시스템을 제안한다. 뉴스 스냅샷은 대규모 뉴스 기사의 텍스트 처리를 바탕으로 이슈의 맥락을 제공하여 독자의 시간을 아끼고 동시에 이슈에 대한 이해를 돕는 것을 목표로 디자인되었다.

본 연구는 다음과 같은 주요 연구 질문들을 바탕으로 진행되었다.

RQ1: 우리의 시스템은 동일한 사건을 보도하는 중복된 뉴스 기사들의 주요 내용을 식별하는 데 도움을 줄 수 있는가?

RQ2: 우리의 시스템은 시간이 지남에 따른 이슈의 주요 내용 변화를 파악하는 데 도움을 줄 수 있는가?

2. 시스템 설계 및 구현 상세

본 시스템은 뉴스 데이터를 처리하는 서버 프로그램과, 사용자와 인터랙션하는 웹인터페이스로 구성된다.

2.1 문장 클러스터링을 활용한 뉴스 시스템 설계

본 연구에서는 뉴스 기사의 문장들을 클러스터링 하는 방안을 고안하였다. 뉴스 기사의 문장들을 클러스터링하면 비슷한 문장들끼리 묶인 군집들을 얻을 수 있으며, 각 군집의 내용들은 전체 뉴스 기사에서 다루고 있는 서브 토픽들로 간주할 수 있다. 이와 같은

방법을 통해 뉴스 기사 그룹의 핵심 내용을 추린다. 또한 여러 날짜에 걸친 이슈의 뉴스 데이터를 클러스터링하고 군집을 시간의 순서로 나열하면 시간에 따른 이슈의 주요 내용 변화를 알 수 있게 된다.

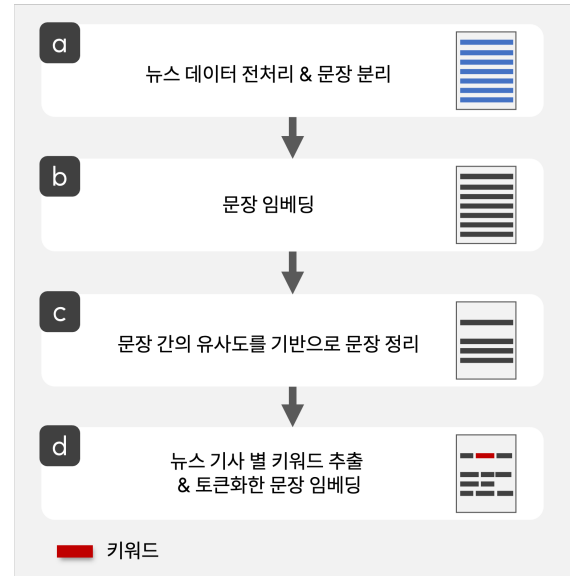


그림 2. 뉴스 데이터 처리 파이프라인

2.2 데이터 수집 및 전처리

2023년 7월 4일부터 10월 20일 까지 네이버 뉴스홈의 헤드라인 기사들을 수집하였다. 수집한 뉴스 데이터는 전처리를 통해 HTML 태그 스타일과 공백, 탭 등을 포함한 특수 문자를 삭제하고, 뉴스 기사 특유의 텍스트 패턴을 제거하였다 (그림 2.a).

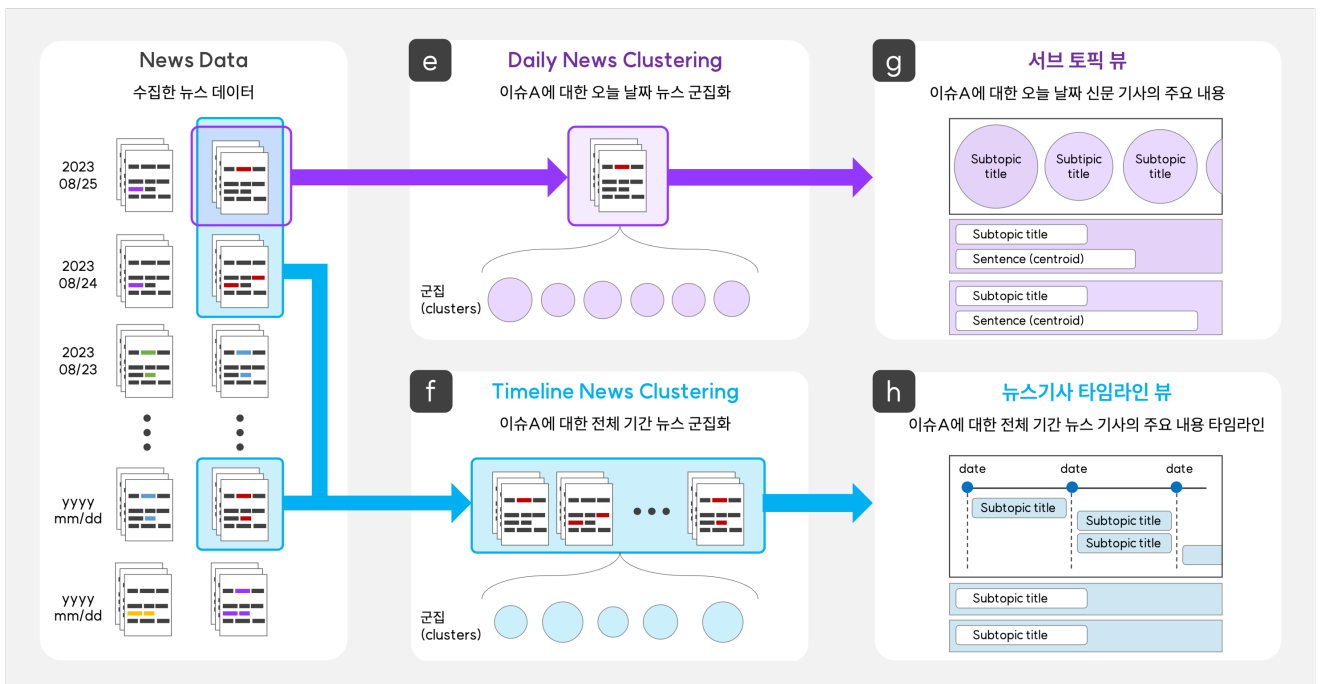


그림 3. 뉴스 데이터의 클러스터링에 따른 사용자 인터페이스 구성: 목적에 따라 두 가지 종류의 클러스터링이 있으며, 각 클러스터링의 결과를 인터페이스에서 시각화하여 전달한다.

2.3 데이터 임베딩 및 클러스터링

■ 문장 임베딩(Sentence Embedding)

수집된 뉴스 기사는 문장 분리 후, Sentence-BERT ('snunlp/KR-SBERT-V40K-klueNLI-augSTS') 모델을 이용해 문장 임베딩을 수행하고 (그림 2.b), 문장 간의 유사도에 따라 불필요한 문장 정리 과정을 거쳤다 (그림 2.c). 이 후, 토큰을 합친 형태로 문장 임베딩을 다시 수행하였다 (그림 2.d).

■ 임베딩한 문장의 클러스터링

클러스터링은 목적에 따라 두 종류가 있다. 먼저, Daily News Clustering 은 오늘 날짜의 뉴스 기사에서 이슈에 대한 주요 내용을 추출하기 위한 것이다. 동일한 날짜의 뉴스 데이터에서 같은 헤드라인 뉴스로 라벨링 된 뉴스 기사의 문장들을 클러스터링함으로써 헤드라인 뉴스 별 하위 주제로 그룹화된 문장 클러스터들을 얻을 수 있다(그림 3.e).

Timeline News Clustering 은 시간의 흐름에 따른 이슈의 주요 내용 변화를 확인하기 위한 것으로, 동일 이슈에 대한 뉴스 기사들을 모아 클러스터링을 수행한다. 그 결과로 이슈 별 하위 주제로 그룹화된 문장 클러스터들을 얻을 수 있으며, 각 클러스터에는 여러 날짜에 걸친 문장들이 포함되어 있다(그림 3.f).

Timeline News Clustering 을 위한 동일 이슈의 뉴스 기사 판별은 Faiss API IndexIDMap 을 활용한 시맨틱 검색과 뉴스 기사 키워드를 활용한다. 뉴스 기사 제목을 쿼리로 시맨틱 검색을 수행하여 결과 문장들이 속한 뉴스 기사를 중복 없이 수집한다. 뉴스 기사의 키워드가 대상 뉴스 기사와 하나 이상 일치하는 경우에만 클러스터링 데이터로 활용한다. 그 이유는 관련 없는 뉴스 기사에서 우연히 언급된 문장을 결과에서 배제하기 위함이다.

클러스터링은 Meta 의 Faiss API 를 이용하였고, optimal K 값은 elbow method 로 자동 할당했다. 각 군집의 중심점(Centroid)에 해당하는 문장이 해당 군집을 대표한다고 가정하였다. GPT-3.5 Turbo 모델에 군집의 대표 문장을 전달하고, 이를 바탕으로 주요 단어를 기반으로 문장의 의미를 함축한 키워드 형태의 Subtopic title 을 생성했다.

3. 뉴스 스냅샷 인터페이스

뉴스 스냅샷 인터페이스는 서브 토픽 뷰(그림 3.g)와 뉴스 기사 타임라인 뷰(그림 3.h)로 구성된다.

3.1 서브 토픽 뷰

Daily News Clustering 을 통해 얻은 각 군집들은 오늘 날짜의 뉴스 기사에서 동일한 이슈를 다루는 뉴스 헤드라인에 대한 서브 토픽을 나타낸다. 이 군집들은 뉴스 스냅샷 인터페이스의 '서브 토픽 뷰'에서 시각화된다. 그림 4 에서 보는 바와 같이, 군집을 시각화 한 부분과 디테일 정보 목록을 함께 제공한다. 문장 군집을 원 형태로 표현하고, 각 군집의 문장 개수 비율에 따라 원의 크기를 설정했다. 목록에는 각 군집의 대표 문장을 함께 표시하였고, 항목을 클릭하면 대표 문장을 추출한 원본 뉴스 기사로 연결된다. 텍스트 맥락을 고려하기 위한 방법으로 군집 내 문장들의 원본 뉴스 기사에서 위치값을 계산하여 이 값을 오름차순 정렬한 순서로 서브 토픽 항목을 전달한다. 비중이 큰 상위 세 개의 항목은 색상으로 강조했다.

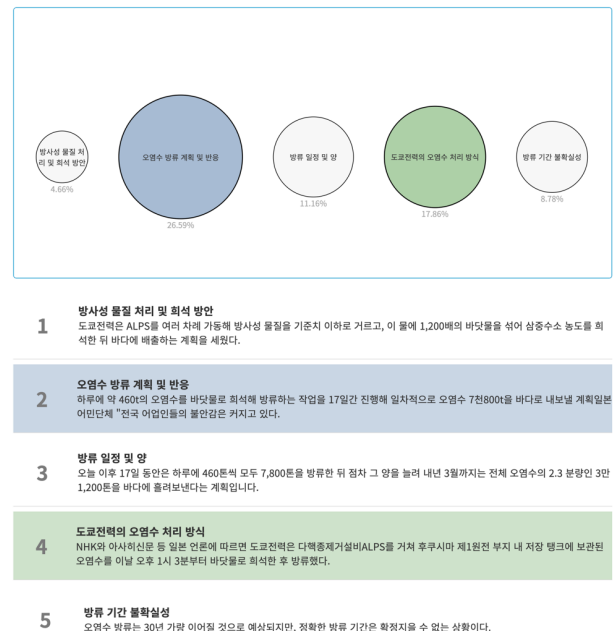


그림 4. 서브 토픽 뷰: 상단에는 군집과 Subtopic title 을 제공한다. 하단에는 Subtopic title 과 순서가 매칭된 대표 문장이 리스트로 제공되어, 뉴스 기사처럼 읽을 수 있다.

3.2 뉴스 기사 타임라인 뷰

Timeline News Clustering 을 통해 얻은 군집들은 전체 기간의 뉴스 기사에서 동일한 이슈에 대한 여러 날짜에 걸친 서브 토픽들을 의미한다. '뉴스 기사 타임라인 뷰'에서는 이 군집들을 시간 순서로 정렬하고 시각화하여 독자가 시간이 지남에 따른 이슈 내용의 변화를 관찰할 수 있도록 한다.

기본 화면인 '심플 타임라인'(그림 5)에는 각 날짜 아래에 이슈의 주요 내용이 짧은 문구로 제공된다.

'다이나믹 타임라인'(그림 6)은 여러 날짜에 걸친 뉴스 기사 그룹 전체에서 군집의 분포를 보여준다. 이를 통해 언제 관련 서브 토픽이 정점에 도달하고 소멸하였는지 트렌드를 파악할 수 있다. 하단에는 각 영역의 대표 뉴스 기사의 링크를 표로 제공한다.

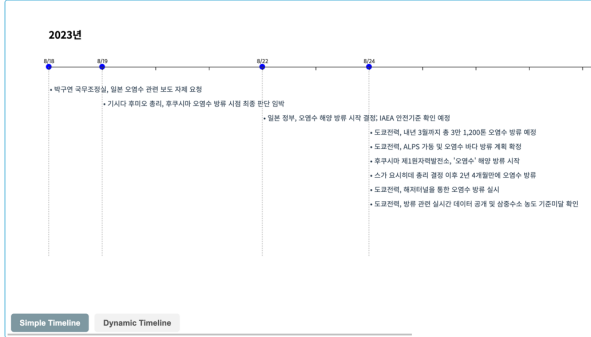


그림 5. 뉴스 기사 타임라인 뷰 '심플 타임라인': 시간이 지남에 따라 이슈의 주요 내용을 보여준다.

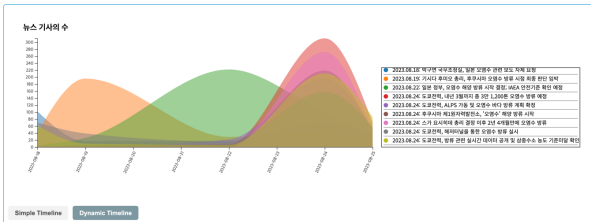


그림 6. 뉴스 기사 타임라인 뷰 '다이나믹 타임라인': 시간이 지남에 따라 각 주요 내용이 언제 얼마나 많이 다루어졌는지 시각화하여 보여준다.

3.3 사용 시나리오

여행사 기획팀에 근무하는 S는 8월 내내 가을 여행 프로모션 준비로 뉴스를 확인하지 못했다. 8월 말, 모처럼 여유가 생긴 S는 오랜만에 '뉴스 스냅샷'에 접속했다. 그는 오늘 날짜(2023.8.25)의 '세계' 뉴스에서 후쿠시마 원전 오염수 해양 방류 개시 관련 뉴스를 탐색했다. '뉴스 기사 타임라인 뷰'를 통해 지난 8월 22일 오염수 해양 방류가 공식 결정되었고, 이를 후인 24일부터 후쿠시마 앞바다에 오염수가 방출되기 시작되었다는 것을 알게 되었다. 그리고 '서브 토픽 뷰'에서 오염수 방류 일정 및 현지의 우려 섞인 반응을 확인했다. 그는 여행 패키지에 포함된 식당 메뉴의 수산물 원산지를 사전 점검하기로 했다. 또한 여행객들에게 안전한 식재료만 제공한다는 사실을 가을 여행 프로모션의 홍보 포인트로 활용하기로 했다.

4. 사용자 평가 및 결과

우리의 시스템이 뉴스 기사의 주요 내용을 식별하고 이슈의 흐름을 파악하는 데 도움을 줄 수 있는지

평가하기 위해 실험을 수행했다. 그 결과를 바탕으로 시스템의 효과 및 사용성을 검토하였다.

4.1 참여자

온라인 커뮤니티 및 눈덩이 표집 방법을 통해 총 22명(P01-P22)의 실험 참여자를 모집하였다. (22명, 남성: N=9, Mean=28.22, SD=5.72, 여성: N=13, Mean=27, SD=6.58)

4.2 실험 디자인 및 태스크

실험은 두 개의 태스크와 사후 인터뷰로 진행되었다. 태스크 1에서는, 하나의 사건에 대한 여러 뉴스 기사들 중에서 원하는 기사들을 선택하여 읽은 후, 화면을 끈 상태에서 사건에 대한 핵심 내용을 주요 키워드로 작성하도록 하였다. 태스크 2는, 시간이 지남에 따라 지속되는 이슈에 대하여 여러 날짜에 걸쳐 발행된 관련 뉴스 기사들 중에서 원하는 뉴스 기사들을 선택하여 읽고, 뉴스 기사 화면을 끈 상태에서 해당 이슈에 대한 핵심 내용을 시간의 흐름에 따라 작성하도록 하였다.

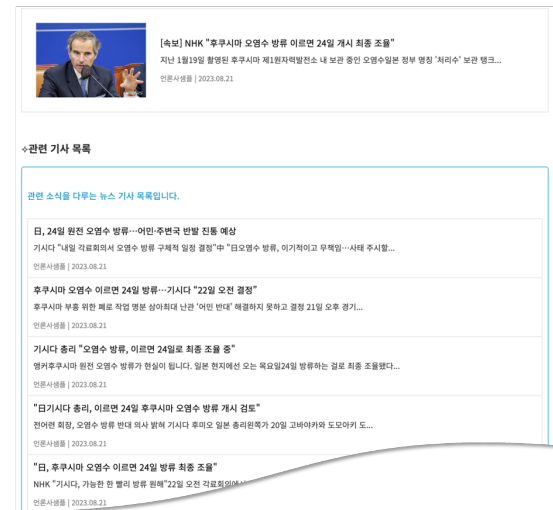


그림 7. 대조군 인터페이스(Control interface): 헤드라인 기사와 동일 이슈를 다루는 뉴스 기사 목록을 제공한다.

제안하는 인터페이스의 효과를 검증하기 위하여 대조군 인터페이스를 제작하였다(그림 7 참조). 대조군 인터페이스는 뉴스 기사를 수집했던 네이버 뉴스홈의 인터페이스를 참고하여, 뉴스 기사 헤드라인마다 대표 뉴스 기사 한 건과 동일한 이슈를 다루는 뉴스 기사의 목록으로 제공된다.

참가자들 간 개인차를 최소화하기 위하여, 실험은 참가자 내 설계(within-subject design)로 진행했으며, 실험 시, 각 인터페이스는 '시스템 A'와 '시스템 B'라는 중립적인 명칭을 사용했다. 실험은 다음과 같이 {(태스크 1-시스템 A), (태스크 1-시스템 B), (태스크 2-시스템 A), (태스크 2-시스템 B)} 네

부분으로 진행되었다. 태스크와 시스템의 순열 별로 서로 다른 뉴스 토픽이 사용되었으며, 실험 전 사전 설문을 통해 조사한 참여자의 사전 지식을 바탕으로 뉴스를 할당하였다. 각 부분이 마칠 때마다 태스크 수행에 대한 전반적인 평가 및 인터페이스의 사용성에 대한 설문을 진행하였다. 설문은 선행 연구[10][13]를 참고하여 구성하였다 (표 1, 표 2 참조).

표 1. 태스크 수행에 대한 전반적인 평가를 위한 질문

Q1. 뉴스 기사 주요 내용 및 주요 내용의 변화를 잘 전달한다.
Q2. 뉴스 기사를 추가적으로 봐야할 필요가 있다.(역문항)
Q3. 어떤 기사를 읽을 지 선택의 어려움이 있다. (역문항)
Q4. 뉴스 기사(이슈) 내용의 이해도를 높인다.
Q5. 뉴스 기사를 읽는 시간을 줄여준다.

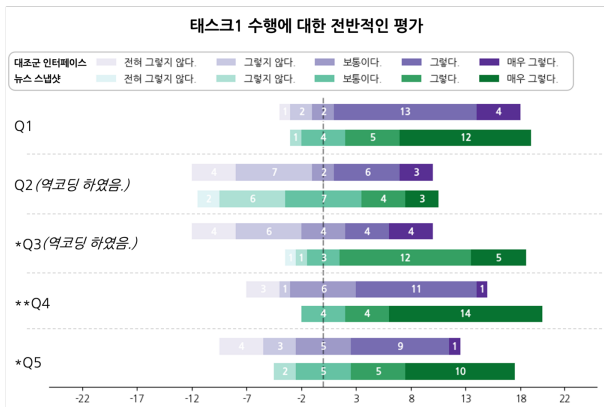


그림 8. 태스크 1 수행에 대한 전반적인 평가 비교: Wilcoxon signed-rank test 결과, *($p < 0.05$), **($p < 0.01$)

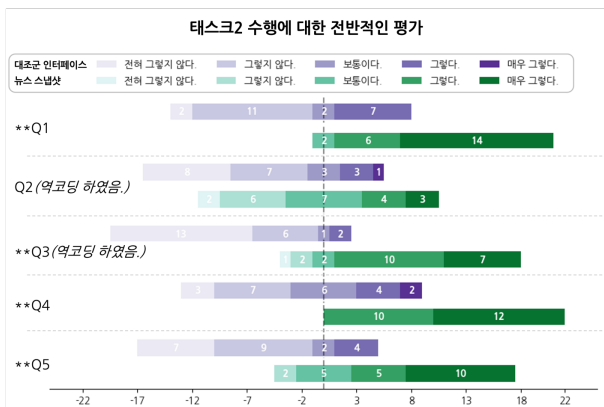


그림 9. 태스크 2 수행에 대한 전반적인 평가 비교: Wilcoxon signed-rank test 결과, **($p < 0.01$)

4.3 결과

태스크 수행에 대한 전반적인 피드백에서 대체로 뉴스 스냅샷에 대한 응답이 더 긍정적인 것으로 드러났다. 그림 8, 그림 9는 태스크 수행에 대한 평가를 그래프로 비교한 것이다. 역문항인 Q2와 Q3은 분석의 편의를 위하여 역코딩하였다. 태스크 1과 태스크 2의 Q1, Q3, Q4, Q5의 경우, 뉴스 스냅샷의 '매우 그렇다' 비율이 대조군 인터페이스보다 훨씬 높다(표 1, 그림 8, 그림 9 참조).

인터페이스 사용성 평가에서는 주로 뉴스 스냅샷에 대한 사용자의 평가가 높게 나타났다. 특히 '뉴스 기사 타임라인 뷰'에 대해 측정한 태스크 2에서 그 효과가 분명하게 드러났다(표 2, 그림 10, 그림 11 참조).

표 2. 태스크 수행에 대한 인터페이스 사용성 비교 질문

유용성	나는 이 인터페이스가 뉴스 기사를 읽고 내용을 파악하는 데 유용하다고 생각한다.
사용 용이성	이 인터페이스는 사용하기 쉽다.
즐거움	이 인터페이스를 사용하는 것이 재미있었다.
효과성	이 인터페이스는 태스크 수행에 효과적이다.
전반적인 만족도	나는 전반적으로 이 시스템에 만족한다.

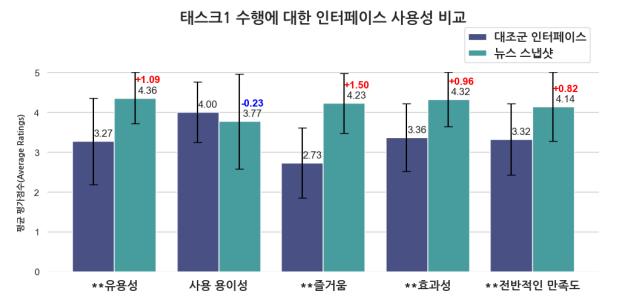


그림 10. 태스크 1 수행에 대한 인터페이스 사용성 평가: Wilcoxon signed rank test 결과, *($p < 0.01$). 에러 막대는 표준 편차를 나타낸다.

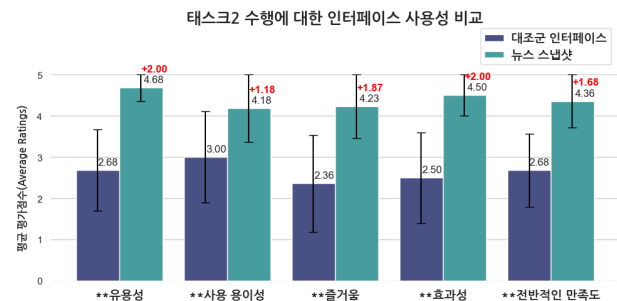


그림 11. 태스크 2 수행에 대한 인터페이스 사용성 평가: Wilcoxon signed rank test 결과, *($p < 0.01$). 에러 막대는 표준 편차를 나타낸다.

■ 사후 인터뷰를 통한 정성적 분석

사후 인터뷰에서 참여자들은 뉴스 스냅샷을 통해 사건의 핵심 내용을 사전에 인지한 상황에서 뉴스 기사를 읽었을 때 이슈의 내용을 보다 수월하게 파악할 수 있었다고 응답하였다(P05, P06, P09). 뉴스 기사의 대표 문장들을 모아둔 목록은 마치 요약된 뉴스 기사 한 부를 읽는 것처럼 느껴져서 유용했다는 의견도 있었다(P01, P03, P06). 문장은 핵심 키워드에 포함되지 못한 자세한 수치 정보와 상황을 설명한다. 대부분 참여자들은 ‘서브 토픽 뷰’에서 항목 별 대표 문장이 함께 제시되어 관련 키워드의 내용을 디테일하게 알 수 있다는 것이 장점이라고 응답했다. 문장의 표현 방식이나 어투는 사실에 대한 미묘한 뉘앙스를 전달하며 감성적인 면에서는 독자에게 친절한 느낌을 전달한다. 참여자(P17)의 경우에는 키워드와 문장이 함께 제시되는 화면이 친절하게 느껴져서 좋았다고 언급했다. ‘뉴스 기사 타임라인 뷰’에 대해서는 대부분의 응답자들이 매우 높은 선호도를 표시하며, 앞으로 계속 사용하고 싶은 기능으로 언급하였다.

5. 논의

이 연구에서는 LLM을 이용하여 뉴스 기사의 많은 내용을 처리하고 뉴스 기사의 맥락을 제공함으로써 독자들의 정보 부담을 줄여주고, 이슈의 주요 내용을 빠르게 파악할 수 있도록 하였다. 뉴스 기사의 타임라인은 이슈가 진행되고 있는 중간 시점에서 뉴스 기사를 접하더라도 이 전의 뉴스 기사 내용을 모두 읽을 필요가 없도록 이슈의 이전 맥락을 제공하여 독자의 이해를 도울 수 있다. 뉴스 기사 타임라인의 시간 축은 뉴스 기사가 발행된 시점과 매칭된다. 이렇게 함으로써 어떤 사실이 언제 밝혀졌고, 그로 인해 상황이 어떤 국면에 접어들게 되었는지와 같이 이슈의 진행 과정 또는 해석 과정을 독자가 파악하는데 도움을 준다.

뉴스 기사의 타임라인을 활용하는 또 다른 방법으로 뉴스 미디어가 이슈를 보도하는 경향이나 사건 보도에 대한 언론의 관심도를 분석하고 이해하는 데 사용될 수 있다. 예를 들어, 스캔들 보도가 발행된 후에 그 이슈에 대한 정정 보도가 발행된 경우, 사건의 보도를 다루는 뉴스 기사 대비 얼마나 많은 뉴스 기사가 적극적으로 정정 보도를 다루는 지 확인할 수 있다.

대부분의 뉴스 기사는 속보나 당일의 사건을 다루는 경우가 많아서 뉴스 기사와 사건의 타임라인 순서가

일치하는 경우가 많다. 이 연구에서는 그러한 뉴스 기사를 대상으로 수행되었다. 만약 사건의 타임라인이 복잡하여 뉴스 기사와 순서가 일치하지 않는 경우에는 뉴스 기사의 텍스트에서 시간 정보를 추출하여, 이슈의 타임라인을 새롭게 구축하는 것이 유용할 수 있으며, 이는 후속 연구의 방향성이 될 수 있다.

6. 결론

이 연구에서는 뉴스 기사의 단편화와 너무 많은 뉴스 기사의 양으로 발생하는 불편함을 완화하기 위하여 ‘뉴스 스냅샷’ 시스템을 제안하고, 그 효과 및 사용성에 대해 검토하였다. 뉴스 스냅샷은 대규모 뉴스 기사 컬렉션에서 동일한 이슈를 다루는 뉴스 기사의 문장들을 분석하고, 이를 두 가지 뷰로 전달한다. 먼저, ‘서브토픽 뷰’에서는 단일 사건에 대한 뉴스 기사의 주요 내용을 제공한다. ‘뉴스 기사 타임라인 뷰’는 지속적인 이슈에 대하여 뉴스 기사의 시간에 따른 주요 내용 변화를 효과적으로 전달한다. 실험 결과, 뉴스 스냅샷은 사용자가 뉴스 기사의 내용을 파악하고 이슈를 이해하는데 도움을 주는 것으로 나타났다. 이 연구에서는 몇 가지 한계점이 있다. 먼저, 한정된 뉴스 기사 데이터셋으로 인해 뉴스 기사의 기한은 최대 3 개월로 제한되었다. 또한, 전처리 및 클러스터링 방법을 더 다양하게 적용하고 테스트했다면 콘텐츠의 품질을 전반적으로 향상시킬 수 있었을 것이라는 아쉬움이 있다. 향후 연구에서는, 이를 바탕으로 뉴스 산업을 발전시키고 독자에게 가치를 전달하는 더 나은 뉴스 시스템 및 서비스에 대한 아이디어를 연구할 것을 제안한다.

참고 문헌

1. Nic Newman, Richard Fletcher, Craig T Robertson, Kirsten Eddy, and Rasmus Kleis Nielsen. Reuters Institute Digital News Report 2023. Reuters Institute. 2023.
2. Nic Newman, Richard Fletcher, Craig T Robertson, Kirsten Eddy, and Rasmus Kleis Nielsen. Reuters Institute Digital News Report 2022. Reuters Institute. 2022.
3. Shin, Y., Sim, H., Jang, Y., Byeon, J., & Yang, J. Media Users in Korea 2022. Korea Press Foundation. 2022.
4. Gurr, G., & Metag, J. What Leads to Audience Issue Fatigue? A Linkage Analysis Study on the Effects of News Coverage on Fatigue From

Ongoing News Issues. *International Journal of Communication*, 17. pp.21. 2023.

5. York C. Overloaded by the news: Effects of news exposure and enjoyment on reporting information overload. *Communication Research Reports*, 30(4), pp.282–292. 2013.
6. O'Brien, H. L. Exploring user engagement in online news interactions. *Proceedings of the American society for information science and technology*, 48(1), pp.1–10. 2011.
7. Lewis, J., & Cushion, S. The thirst to be first: An analysis of breaking news stories and their impact on the quality of 24-hour news coverage in the UK. *Journalism Practice*, 3(3), pp.304–318. 2009.
8. Doshi, K., Gokhale, S., Mamtara, H., & Bide, P. Analytics and visualization of trends in news articles. In 2019 ICAC3. 2019.
9. Yuping, H., Ligu, Z., Lei, Z., & Ruisong, Z. NDVS: A System of News Data Visualize. In 2015 IMCCC, pp. 475–479. 2015.
10. Gao, M., Do, H. J., & Fu, W. T. Burst your bubble! an intelligent system for improving awareness of diverse social opinions. In 23rd ACM IUI. pp.371–383. 2018.
11. Munson, S., Lee, S., & Resnick, P. Encouraging reading of diverse political viewpoints with a browser widget. In *Proceedings of ICWSM*. Vol. 7, No. 1, pp. 419–428. 2013.
12. Park, S., Kang, S., Chung, S., & Song, J. NewsCube: delivering multiple aspects of news to mitigate media bias. In *Proceedings of the SIGCHI conference on human factors in computing systems*. pp. 443–452. 2009.
13. Mishra, A., Ginpalli, S., & Bryan, C. News Kaleidoscope: Visual Investigation of Coverage Diversity in News Event Reporting. In 2022 IEEE 15th PacificVis. pp.131–140. 2022.
14. Sultanum, N., Bezerianos, A., & Chevalier, F. Text visualization and close reading for journalism with storifier. In 2021 IEEE VIS. pp.186–190. 2021.