



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

석사학위논문

인공지능과 함께 읽기

: 한국 근현대 시(詩) 감정 데이터 색채 시각화 연구

한 국 학 중 앙 연 구 원

한 국 학 대 학 원

인문정보학전공 : 임이로

석사학위논문

인공지능과 함께 읽기

: 한국 근현대 시(詩) 감정 데이터 색채 시각화 연구

지도교수 김 병 준

한 국 학 중 앙 연 구 원

한 국 학 대 학 원

인문정보학전공 : 임 이 로

인공지능과 함께 읽기

: 한국 근현대 시(詩) 감정 데이터 색채 시각화 연구

석사학위 논문으로 제출함.

(2026. 05. 31.)

한 국 학 중 앙 연 구 원

한 국 학 대 학 원

인문정보학전공 : 임이로

이 논문을 임이로의 문학석사 학위논문으로 인정함.

2026년 06월 일

심사위원장 서동수

심사위원 김바로

심사위원 김병준

한국학중앙연구원

한국학대학원

목 차

국문초록	vi
I. 서론	1
1. 연구 배경 및 목적	1
2. 이론적 배경	3
1) 문학 데이터셋의 환경적 전환	3
2) 시의 감정과 데이터	7
3) 시와 색채	12
3. 연구 방법론	15
II. 함께 읽기를 위한 데이터셋 및 모델 구축	21
1. 한국 근현대 시 감정 데이터셋 KPoEM 구축	21
1) KPoEM 데이터셋 개괄	21
2) 데이터셋 구축 과정	27
3) 데이터셋 정량 분석 및 평가	33
2. KPoEM 감정 분류 모델 개발	35
3. 한국인 색채 감정 데이터셋 KCoEM 구축	40
1) KCoEM 데이터셋 개괄	40
2) 데이터셋 구축 과정	44
3) 데이터셋 정량 분석 및 평가	56
4. 감정-색채 변환 알고리즘 개발	58
1) 감정 속성 기반 배색 방식 결정	58
2) 주조색 및 보조색 결정	59
3) 색상 결정 예외처리	61

4) 색조 조정 규칙 설정	62
5) 2색 배색 팔레트 면적비 구성	63
III. 함께 읽기(Co-Reading) 시스템 및 웹 인터페이스 구현	64
1. 함께 읽기 시스템의 커뮤니케이션 구조 설계	64
2. 함께 읽기 시스템 파이프라인 설계	66
1) 감정 분류 및 배색 방식 결정	67
2) 주조색 및 보조색 결정	68
3) 컨텍스트 엔지니어링 기반 이미지 형용사 선택	69
4) 2색 배색 팔레트 선택	74
5) 색조 조정 및 시각화	77
3. 웹 인터페이스 구현 및 결과	81
1) 가까이 읽기 인터페이스	83
2) 멀리서 읽기 인터페이스	86
3) 함께 읽기 인터페이스	91
4) 가까이 읽기, 멀리서 읽기, 함께 읽기 인터페이스 비교	94
IV. 결론	96
참고문헌	99
Abstract	109
부록 1. KOTE 감정 어휘와 KCoEM 감정 어휘 대응 관계	111

표 목 차

표 I -1. KOTE와 KENT의 감정분류체계 상관관계	9
표 I -2. 2색 배색 감성 평가 3요소	14
표 II -1a. KPoEM 행 단위 데이터셋 예시	22
표 II -1b. KPoEM 작품 단위 데이터셋 예시	23
표 II -2a. KPoEM 행 단위 데이터셋 메타데이터 스키마	25
표 II -2b. KPoEM 작품 단위 데이터셋 메타데이터 스키마	25
표 II -3. KPoEM 데이터셋 주석자 구성	26
표 II -4. KOTE 44개 감정 어휘와 인덱스(id)	26
표 II -5. 한용운, 「당신을 보았습니다」 JSON 데이터 예시	28
표 II -6. 한용운, 「당신을 보았습니다」 행 단위 분절 데이터 예시	29
표 II -7. 한용운, 「타골의 시를 읽고」 한글-한자 병기 과정	30
표 II -8a. 이상, 「꽃나무」 원문 텍스트	31
표 II -8b. 이상, 「꽃나무」 문장 분절 예외 처리	31
표 II -9. KPoEM 데이터셋 통계	33
표 II -10. KPoEM 데이터셋 빈도 상위 10개 감정 어휘	33
표 II -11. 주석자 수에 따른 감정 주석 일치 통계	34
표 II -12. KPoEM 데이터셋 학습·검증·테스트 세트 분할	35
표 II -13. 감정 분류 모델 성능 평가 비교 (Threshold=0.3)	36
표 II -14. 감정 분류 모델 정성 평가 비교	37
표 II -15. KCoEM 데이터셋 예시	42
표 II -16. KCoEM 데이터셋 메타데이터 스키마	43
표 II -17. 한지운과 나건(2021)의 18개 감정 어휘	45
표 II -18. 김애경과 오윤경(2016)의 긍정·부정 어휘	46
표 II -19. 김옥(2017)의 기본 감정 및 세부 감정 범주	47
표 II -20. 정현희(2018)의 문헌 중 KCoEM 구축에 활용한 항목	47

표 II-21. 문재호와 신진우(2021)의 색에 따른 감정 효과 자료 발췌	48
표 II-22. KCoEM 구축에 활용한 Nijdam(2006) 연구의 감정 어휘	49
표 II-23. KS 색체계 계통색 이름과 계통색 약호	51
표 II-24. 한지운과 나건(2021)의 실험 결과 예시	52
표 II-25. KS 색체계와 PCCS 색체계 비교	52
표 II-26. 김애경과 오윤경(2016)의 실험 결과 예시 (PCCS 색체계)	53
표 II-27. 김옥(2017)의 실험 결과 예시 (먼셀/KS 색체계)	53
표 II-28. Naz Kaya의 색채-감정 모델 예시 (먼셀 색체계)	53
표 II-29. 선행 문헌에서 색상 이름만 제시한 경우 RGB 값 변환 예시	54
표 II-30. 문헌 유형별 우선순위	54
표 II-31. KCoEM 데이터셋 우선순위 예시	55
표 II-32. KCoEM 데이터셋 출처 유형별 구성 통계	56
표 II-33. KCoEM 데이터셋 색상 범주별 분포 통계	57
표 II-34. KOTE가 제시한 감정의 긍정·부정·중립 분류	58
표 II-35. KOTE 감정 속성에 따른 배색 방식	58
표 II-36. 보조색 결정 우선순위	60
표 III-1. 신석정, 「네 눈망울에서는」의 감정 분류 결과	67
표 III-2. 신석정 사례의 주 감정에 대한 KCoEM 데이터	68
표 III-3. 신석정 사례의 보조 감정에 대한 KCoEM 데이터	69
표 III-4. 정제한 I.R.I 이미지 형용사	71
표 III-5. 이미지 형용사 선택을 위한 LLM 프롬프트 구성	72
표 III-6. 신석정 사례의 감정 분류와 이미지 형용사 선택 결과	74
표 III-7. ‘온화한’ 형용사군의 배색 팔레트 데이터 예시	76
표 III-8. 신석정 사례의 주조색과 보조색 색조 조정 결과	77
표 III-9a. 백석(2023), 「나와 나타샤와 흰 당나귀」 함께 읽기 결과	79
표 III-9b. 김수영(1974), 「푸른 하늘을」 함께 읽기 결과	80
표 III-10. 가까이 읽기, 멀리서 읽기, 함께 읽기 인터페이스 비교	94

그 림 목 차

그림 I -1. 연구 프로세스 개념도	15
그림 II-1. 위키문헌에 수록된 한용운, 「당신을 보았습니다」	28
그림 II-2. 먼셀(Munsell)의 10색상환	60
그림 III-1. 함께 읽기(Co-Reading)의 커뮤니케이션 구조	64
그림 III-2. 함께 읽기(Co-Reading) 시스템 파이프라인 설계도	66
그림 III-3. I.R.I 형용사 이미지 스케일	70
그림 III-4. I.R.I 배색 이미지 스케일의 배색 그룹 및 색상 ID 체계	75
그림 III-5. 신석정, 「네 눈망울에서는」 최종 시각화 2색 배색 팔레트	78
그림 III-6. KPoEM 인터페이스 프로토타입 구조 설계 개념도	81
그림 III-7. 김소월, 「진달래꽃」 가까이 읽기 인터페이스 예시	83
그림 III-8. 김소월 전체 작품의 멀리서 읽기 인터페이스 예시	86
그림 III-9. 김소월 작품의 SHAP 기반 워드클라우드	88
그림 III-10. 김소월 작품 감정 분포 UI	89
그림 III-11. 김소월 작품-감정 히트맵 UI 예시	90
그림 III-12. 김소월, 「진달래꽃」 함께 읽기 인터페이스 예시	91

국문초록

인공지능과 함께 읽기 : 한국 근현대 시(詩) 감정 데이터 색채 시각화 연구¹⁾

임 이 로

본 연구는 인간과 인공지능이 문학 텍스트를 함께 해석하는 환경, ‘함께 읽기 (Co-Reading)’ 시스템을 제안한다. 이를 위해 본 연구는 데이터셋을 인간과 인공지능의 해석을 매개하는 환경적 미디어(environmental media)로 조망하며, 인공지능을 인간과 함께 의미를 구성하는 공동 독자(co-reader)로 재정의했다. 이러한 해석 환경을 구현하기 위해 본 연구는 해석의 다의성이 두드러지는 문학, 시(詩)를 대상으로 선정했다. 또한, 색채가 시에서 감정과 심상을 매개하는 표현 요소라는 점에 주목하여, 시의 감정을 색채로 시각화하는 인간-인공지능 ‘함께 읽기’ 시스템을 설계하고 개발했다. 연구 과정은 다음과 같이 진행했다.

첫째, 한국 근현대 시를 대상으로 인간의 감정 해석을 구조화한 시 감정 데이터셋 KPoEM(Korean Poetry Emotion Mapping)을 구축했다. 그리고 이를 기반으로 시 텍스트의 감정을 분류하는 KPoEM 감정 분류 모델을 개발했다.

둘째, 기존 색채 심리 연구를 바탕으로 한국인 색채 감정 데이터셋 KCoEM(Korean Color Emotion Mapping)을 구축하고, 감정의 속성과 배색 원리를 반영한 감정-색채 변환 알고리즘을 개발했다.

1) 이 논문은 2025년 대한민국 교육부와 한국연구재단의 인문사회분야 석사과정생연구장려금 지원사업의 지원을 받아 수행된 연구임(NRF-2025S1A5B5A20019820).

또한 본 논문은 필자가 참여한 공동연구의 성과인 KPoEM 데이터셋 구축 및 감정 분류 모델 개발 내용을 기반 자료로 활용했다. 관련 내용은 다음 연구에 기반한다.

Lim, I., Ji, H., & Kim, B. (2026). KPoEM: A human-annotated dataset for emotion classification and RAG-based poetry generation in Korean modern poetry. *The Review of Korean Studies*, 29(1), 161-206.

<https://doi.org/10.25024/review.2026.29.1.006>

나아가 본 논문에서 제안한 KCoEM 데이터셋 구축, 감정-색채 변환 알고리즘, 함께 읽기 (Co-Reading), 그리고 웹 인터페이스 설계 및 해석 구조는 필자의 독자적인 연구 수행 결과에 해당한다.

셋째, 나아가 컨텍스트 엔지니어링을 적용하여 입력된 시에 대한 인간의 감정과 인공지능의 해석 결과를 2색 배색 팔레트로 재구성하는 함께 읽기(Co-Reading) 시스템을 설계하고, 이를 구현하기 위한 커뮤니케이션 구조와 파이프라인을 구축했다. 넷째, KPoEM 데이터셋을 대상으로 가까이 읽기, 멀리서 읽기, 그리고 인간-인공지능 상호작용에 기반한 함께 읽기를 각각 독립적인 읽기 환경으로 구현하고, 이를 하나의 웹 인터페이스 프로토타입 안에서 비교·탐색할 수 있도록 통합했다. 가까이 읽기와 멀리서 읽기가 시 텍스트에 대한 미시적·거시적 관점을 제공하는 데 집중했다면, ‘함께 읽기’는 인간과 인공지능이 재구성한 색채 이미지를 매개로 인간의 감각적 재해석을 유도한다. 이를 통해 본 연구는 데이터 시각화가 인간과 인공지능이 함께 해석하고 의미를 구성할 수 있는 커뮤니케이션 환경으로 기능할 수 있음을 시사한다.

결과적으로 본 연구는 데이터셋과 인터페이스를 매개로 인간과 인공지능이 동일한 문학 텍스트를 함께 읽고 새로운 의미를 구성하는 디지털 환경을 실험적으로 구현함으로써, 인공지능 시대 포스트쓰기체계(postwriting)의 한 가능성을 제안한다.

주제어 : 함께 읽기, 인간-인공지능 상호작용, 환경적 미디어, 포스트쓰기체계, 감정 데이터셋, 감정-색채 변환, 멀티모달, 데이터 시각화, 웹 인터페이스, 감성 컴퓨팅, 한국 근현대 시, 디지털 인문학

I. 서론

1. 연구 배경 및 목적

지금까지 생성형 인공지능 창작 연구는 ‘인공지능이 무엇을 생성할 수 있는가’에 초점을 맞춰 왔다(Anantrasirichai & Bull, 2021). 이러한 추세에서 우리는 대량의 데이터를 인공지능에 학습시키며, 창의성의 흔적을 찾고자 한다(Ismayilzada et al., 2025). 그러나 이 과정에서 인공지능은 인간의 체험과 맥락으로부터 단절되고, 인간 또한 그 과정에서 해석 주체로서의 위치를 상실한다.

이에 본 연구는 이러한 생성 중심 접근에서 벗어나, 인간과 인공지능이 동일한 문학 텍스트를 대상으로 상호작용하며 의미를 재구성하는 ‘함께 읽기’의 가능성을 탐색한다. 즉, 인공지능을 인간과 함께 문학을 읽고 해석하는 공동 독자(co-reader)로 재위치시키고, 인간과 인공지능의 해석이 교차하는 커뮤니케이션 환경을 설계하는 것을 목표로 한다.

인간과 인공지능이 동일한 텍스트를 함께 해석하기 위해서는 양자를 연결하는 매개 환경이 필요하다. 따라서 본 연구는 ‘데이터셋’을 단순히 학습 자원이 아닌 인간과 인공지능 간 해석을 매개하는 ‘미디어’로 조망한다. 즉, 데이터셋은 객관적 사실의 집합이 아니라 인간의 감정, 심상, 해석이 축적된 환경적 매체이며, 이러한 환경을 기반으로 인간과 인공지능이 서로 의미를 교환하는 시스템을 설계할 수 있다.

문학 텍스트의 의미는 텍스트 내부에 고정되어 존재하는 것이 아니라, 독자의 읽기 과정 속에서 구체화된(Iser, 1980). 특히 시는 언어를 통해 독자마다 서로 다른 감각적 경험을 생성하는 문학 장르다. 그래서 시는 인간의 주관적 경험과 인공지능의 해석이 교차하는 지점을 탐색하는 논의의 출발점으로 적합하다. 또한, 색채는 시의 감정과 심상을 구체화하는 중요한 표현 요소로 기능한다. 따라서 본 연구는 시의 감정과 심상을 색채를 통해 감각적으로 재구성하고, 이를 인간과 인공지능의 해석 결과를 연결하는 매개로 삼고자 한다.

구체적으로 본 연구는 한국 근현대 시 감정 데이터셋, KPoEM(Korean Poetry Emotion Mapping)과 한국인 색채 감정 데이터셋, KCoEM(Korean Color Emotion

Mapping)을 구축했다. 이를 바탕으로 감정 분류 모델, 감정-색채 변환 알고리즘, 그리고 컨텍스트 엔지니어링을 통합하여 인간과 인공지능이 시 텍스트의 감정을 색채 이미지로 재구성하는 함께 읽기(Co-Reading) 시스템을 구현했다.

이를 통해 시의 감정은 언어에서 데이터로, 데이터에서 색채 이미지로 확장된다. 이어서 인간은 그 결과를 다시 탐색하고 해석함으로써 의미 생성 과정에 참여하게 된다. 따라서 함께 읽기 시스템 역시 단일한 결과를 제시하는 데 머무르지 않고, 입력된 시 텍스트에 따라 색채와 의미를 동적으로 재구성하는 읽기 환경을 지향한다.

이러한 다각적 상호작용의 결과는 본 연구가 구축한 데이터셋과 웹 인터페이스 환경을 통해 구현된다. 특히 본 연구에서는 함께 읽기의 특성을 비교하고 검토하기 위해, 전통적인 문학 연구 방법론인 가까이 읽기와 디지털 인문학의 멀리서 읽기, 그리고 본 연구가 제안하는 ‘함께 읽기’를 각각 독립적인 읽기 환경으로 구현하고 하나의 웹 인터페이스 프로토타입으로 통합했다. 이는 정량성과 효율성 중심 데이터 시각화의 한계를 넘어, 인문 데이터 시각화에 인간과 인공지능을 참여시켜 공동 해석과 감각적 감응(感應)이 이루어지는 웹 환경을 구현하려는 시도다. 결과적으로 본 연구는 인간과 인공지능이 공동으로 해석하고 의미를 구성하는 해석 중심의 디지털 매체 환경을 제안한다.

본 연구는 다음 연구 질문을 설정했다.

- RQ1. 인공지능은 인간과 함께 문학 텍스트를 읽고 해석하는 공동 독자(co-reader)로 기능할 수 있는가?
- RQ2. 문학 텍스트의 감정과 심상은 어떠한 방식으로 시각화할 수 있는가?

2. 이론적 배경

1) 문학 데이터셋의 환경적 전환

(1) 미디어로서의 데이터셋

본 연구는 데이터셋을 고정된 정보의 집합이나 기계 학습을 위한 입력 자원으로 간주하는 관점에서 벗어나, 데이터셋 자체를 하나의 환경적 미디어로 조망하고자 한다. McLuhan(1964, p. 7)은 “미디어는 메시지다”라는 명제를 통해, 미디어의 형식 자체가 의미 생산을 위한 핵심 조건임을 강조했다. 이 관점에서 미디어는 정보를 전달하는 수단에서 나아가, 인간의 지각 방식과 사고 구조를 재편하는 환경적 장치로 이해된다. 이러한 논의는 Peters(2015)가 제시한 미디어를 환경으로 이해하는 관점(media as environmental)으로 확장되며, 나아가 인간 존재의 조건을 구성하는 인프라로서 미디어를 인식하는 방향으로 발전한다. Peters에 따르면 미디어는 인간이 세계를 인식하고 관계 맺는 방식 자체를 가능하게 하는 환경이다.

뉴미디어 이론가 Manovich(2001)는 데이터베이스(database)를 사용자 경험과 문화적 인식을 조직하는 새로운 ‘문화적 형식(cultural form)’으로 보았다. 이어서 그는 데이터가 사진이나 영화와 마찬가지로, 고유한 행동 유도성과 제약을 지닌 독립적 매체(medium)라고 설명한다(Manovich, 2020). 즉, 데이터는 인간이 무엇을 재현하고 어떠한 방식으로 사고할 것인지를 구조화하는 매개적 조건으로 작동한다. 이는 데이터가 본질적으로 주어진(given) 것이 아니라 인간의 선택과 해석에 의해 구성된 것, 즉 ‘capta’라는 Drucker(2014)의 관점으로 이어진다. 데이터는 관찰자로부터 독립적인 객관적 사실이 아니라, 특정한 관점과 매개 과정을 통해 형성된 결과물이다. 나아가, D’Ignazio와 Klein(2020)은 데이터가 관점과 가치, 그리고 권력 관계를 반영한 산물임을 강조한다.²⁾ 이와 유사하게, ImageNet 프로젝트³⁾를 둘러싼 논쟁은 데이터셋이 결코 중립적이지 않으며, 분류 체계, 명명 방식, 수집 과정에서 기술적

2) D’Ignazio와 Klein(2020)의 관점은 실제 인공지능 시스템 사례에서도 확인할 수 있다. 예컨대 아마존의 기존 합격자 이력서 데이터를 학습시킨 채용 알고리즘은 남성 지원자를 선호하는 경향을 보였으며, ‘women’이라는 단어가 포함된 이력서나 여성 대학 출신 지원자를 낮게 평가하는 경향을 보였다. 이는 데이터셋이 특정한 사회적 편향과 역사적 조건을 반영하며, 인공지능의 판단 기준 형성에 영향을 미칠 수 있음을 보여준다.

3) ImageNet은 WordNet의 의미 계층 구조를 바탕으로 구축된 대표적인 대규모 이미지 데이터셋 프로젝트다(Deng et al., 2009).

결함뿐만 아니라, 사회적 편향과 권력 구조를 내재해 있음을 명확히 드러낸다 (Crawford, 2021; Kisel et al., 2025). 이러한 논의는 데이터셋이 인공지능의 인식 관점을 형성하며, 그 자체로 인간과 인공지능의 해석 가능성을 조직하는 환경적 미디어로 기능할 수 있음을 시사한다.

즉, 인간은 자신의 해석과 표현을 데이터로 변환하여 데이터셋을 구성하고, 인공지능은 이 데이터셋을 기반으로 세계를 인식하고 해석한다. 이 과정은 한 방향에 그치지 않는다. 인간의 데이터셋을 통해 형성된 인공지능의 관점은 다시 인간의 인식과 판단, 나아가 행동에 영향을 미치고 있다. 따라서 인간과 인공지능은 데이터셋을 매개로 상호 연결된 커뮤니케이션 구조 속에 위치한다.

이에 본 연구는 데이터셋을 인간의 해석과 경험이 축적된 환경적 미디어로 조망한다. 나아가 이를 바탕으로 인간과 인공지능이 동일한 문학 텍스트를 공유하고, 각자의 해석을 상호 참조하여 새로운 의미를 구성하는 함께 읽기(Co-Reading) 시스템을 제안한다.

(2) 문학 텍스트 데이터의 변용

문학 데이터는 다양한 방식으로 구조화되어 새로운 매체 형식으로 재구성될 수 있다. 이러한 과정에서 문학 텍스트는 문자 중심의 표현을 넘어 시각, 청각, 인터페이스 등 다양한 감각적 층위와 결합하게 된다. Goldsmith(2023)가 『문예 비창작』에서 제안한 ‘포스트쓰기체계(postwriting)’ 개념은 이와 같은 디지털 시대의 텍스트 변환 과정에 개념적 근거를 제공한다.⁴⁾

전성규(2024)의 연구는 이러한 관점에서 이광수의 작품을 8만여 개 행으로 분절하고, 이를 프로그래밍이 가능한 데이터 형태로 변환하여 색채 매핑과 사운드 매핑을 시도했다. 사운드 매핑에서는 소리와 관련된 표현을 추출하여 장소성과 서사를 반영한 오디오 텍스트를 생성하고 이를 지도 인터페이스로 시각화했다. 또한, 색채 매핑을 통해 이광수 문학 텍스트에 나타난 색깔 표현을 수집하여 붉은색(욕망), 노란빛(여성, 조선) 등의 의미를 도출하고 작품 속 색의 변화 과정을 분석했다. 이와 같은

4) 포스트쓰기체계(postwriting)는 표절, 재편집, 프로그래밍 등 비창조적 글쓰기의 다양한 실천을 포함하며, 디지털 환경에서 텍스트가 생성되고 변형되는 과정을 새로운 글쓰기의 실천으로 바라보는 접근이다(Goldsmith, 2023).

시도는 문학 텍스트를 색채와 시각적 인터페이스 등 다양한 매체와 결합하는 흐름을 보여준다. 이는 텍스트 중심의 문학을 공감각적 경험의 차원으로 확장하는 ‘탈문자화’ 과정으로, 문학 데이터가 다양한 멀티모달 형식으로 재구성될 수 있음을 시사한다.

본 연구가 제안하는 함께 읽기(Co-Reading) 시스템 역시 시 텍스트의 감정을 데이터화하고, 이를 색채 이미지로 재구성하여 디지털 환경에서 ‘읽기’와 ‘쓰기’의 경계를 흐리게 한다. 이러한 변용 과정은 포스트쓰기체계(postwriting)의 관점에서 새로운 의미 생산 과정으로 이해할 수 있다.

(3) 인문 데이터 시각화 인터페이스

Drucker(2014)는 인문학의 본질이 해석과 모호성, 추론, 그리고 질적 판단에 있다는 점을 강조하며, 자연과학에서 유래한 기존의 막대 차트와 같은 정보 시각화가 인문 데이터 시각화에 과연 적절한지 지적한다. 이어서 그는 시각화 인터페이스가 정보 전달용 도구를 넘어, 사용자의 사고방식과 담론의 방향을 형성하는 주체적인 ‘매개 공간’이 되어야 한다고 했다. 마찬가지로 Manovich(2001)는 데이터가 인터페이스를 통해 사용자와 상호작용하며 다양한 방식으로 탐색될 수 있다는 점을 주목하고, 인터페이스가 의미 생산의 핵심적인 조건으로 작동한다고 보았다.

Poem Viewer와 Poemage는 시 텍스트를 탐색할 수 있는 인터페이스로써, 문학 연구의 ‘가까이 읽기(close reading)’와 ‘멀리서 읽기(distant reading)’를 연결하는 사용자 상호작용 기반 데이터 시각화를 구현했다(Abdul-Rahman et al., 2013; McCurdy et al., 2016).⁵⁾ Jänicke 외(2015) 역시, 데이터 시각화 및 인터랙티브 분석 환경은 사용자가 대규모 데이터 분석과 개별 텍스트 해석을 유기적으로 오갈 수 있도록 지원한다고 설명한다. 이러한 인터랙티브 시각화 인터페이스는 데이터 기반 분석과 인간 해석 사이의 순환적 관계를 강화하며, 기계가 패턴을 드러내고 인간이 이를 해석하는 상호작용적 읽기 구조를 가능하게 한다.

5) 가까이 읽기는 개별 텍스트의 언어적 표현, 수사적 장치, 의미 구조를 정밀하게 분석하는 전통적인 문학 연구 방법으로, 텍스트 내부의 해석 가능성을 심층적으로 탐색하는 데 강점을 지닌다. 반면, 디지털 인문학의 발전과 함께 등장한 멀리서 읽기는 Moretti(2013)에 의해 제안된 개념으로, 대규모 텍스트 데이터를 집합적으로 분석하여 문학 작품의 구조적 패턴과 경향 등을 파악하는 접근이다.

이러한 연구 흐름은 인간과 인공지능이 문학 텍스트를 함께 해석하는 새로운 읽기 환경으로도 확장될 수 있다. 따라서 본 연구는 KPoEM 데이터셋(Lim et al., 2026)을 기반으로 가까이 읽기, 멀리서 읽기, 그리고 함께 읽기의 데이터 시각화 결과를 통합한 인터페이스 프로토타입을 설계했다. 이를 통해 사용자는 서로 다른 읽기 방식의 특징을 비교·탐색할 수 있으며, 데이터 시각화 인터페이스를 해석과 상호작용이 이루어지는 매개 공간으로 경험하게 된다.

(4) 생성형 인공지능의 창의적 활용과 공동 독자(co-reader)

생성형 인공지능에서 나타나는 할루시네이션(hallucination)⁶⁾ 현상은 법률, 의료, 학술 분야와 같이 사실성과 정확성이 중요한 영역에서는 억제되어야 할 오류로 간주된다. 그러나 최근 일부 연구와 언론에서는 이러한 현상을 오류와 실패가 아니라, 창의적 사고와 해석을 확장하는 요소로 재조명하고 있다. Thornhill(2024)은 생성형 인공지능의 할루시네이션 현상이 인간의 창의성을 자극할 수 있다고 설명했다. 특히 Puchner(n.d.)의 사례는 역사적 인물과 철학자를 기반으로 만든 맞춤형 대화봇을 통해, 생성형 인공지능이 인간의 철학적·비판적 사유를 확장하는 상호작용 파트너로 작동할 수 있음을 보여준다.⁷⁾

또한, Wan 외(2024)는 글쓰기 초기 구상 단계에서 LLM을 활용할 경우, 인공지능의 기술적 한계나 예측 불가능성이 오히려 창의성을 촉발하는 요소로 작용할 수 있음을 확인했다. 이들은 인공지능이 산출한 무작위성과 예상치 못한 변이가 사용자의 새로운 발상과 해석을 유도할 수 있다고 주장한다.

문학 텍스트의 의미는 텍스트 내부에 고정된 형태로 존재하기보다 독자의 능동적인 읽기 행위 속에서 구성된다(Iser, 1980). 즉, 똑같은 시어와 표현이라도 독자의 경험과 맥락에 따라 서로 다른 감정과 심상을 형성할 수 있다. 이러한 수용미학적 관점에서 볼 때, 인공지능이 생성하는 예기치 않은 감정 해석과 이미지적 연상 역시, 오

6) 환각 또는 할루시네이션(hallucination)은 생성형 인공지능이 실제 사실이나 논리적 근거와 일치하지 않는 정보를 높은 확신을 가지고 생성하는 현상을 의미한다(위키백과, n.d.).

7) Puchner의 역사적 인물 채팅봇(Chatterbot) 사이트의 주소는 다음과 같다.
Puchner, M. (n.d.). *Custom GPTs and online education*.
<https://www.martinpuchner.com/custom-gpts-and-online-education.html>

답이나 오류로만 치부할 수 없다. 오히려 이는 인간 독자가 미처 주목하지 못한 연관성이나 낯선 해석을 드러내며, 문학 텍스트에 대한 새로운 이해를 촉진할 수 있다.

따라서 본 연구는 인공지능의 예측 불가능한 해석 결과를 문학 텍스트에 대한 잠재적 해석 자원으로 이해하고자 한다. 본 연구가 제안하는 함께 읽기(Co-Reading) 시스템 역시, 시-감정-색채가 연결되어 변환되는 과정에서 발생하는 인공지능의 해석 다양성과 맥락 확장을 수용한다. 이를 통해 인공지능은 단일한 정답을 산출하는 생성 도구가 아니라, 인간과 함께 시의 정서와 심상을 탐색하고 이를 색채 이미지로 재구성하는 ‘공동 독자(co-reader)’로 기능한다.

2) 시의 감정과 데이터

(1) 시의 심상과 감정

시는 언어의 예술이며, 언어 그 자체만으로는 형용할 수 없는 정서적 경험을 불러일으킨다. Day-Lewis(1984)는 시를 여러 심상이 결합하여 하나의 이미지(image)를 형성하는 예술 형식으로 이해했으며, 오규원(2017)은 시의 이미지가 언어의 의미 전달을 넘어 감각적·표상적 경험을 환기하는 표현 방식이라고 보았다. 즉, 시의 이미지는 비유와 상징, 암시와 같은 표현을 통해 독자의 감각과 상상을 자극하며 언어를 감각적 경험으로 전환하는 역할을 한다.

또한 시는 심상을 통해 독자에게 감정을 유발한다. 박민수(1986, p. 124)가 소개한 김억의 시론에 따르면, 김억은 시가(詩歌)를 “이지(理智)를 떠나 감정 세계를 소요하는 것”으로 설명했다.⁸⁾ 이처럼 시에서 감정은 논리적 진술로 표명되기보다는, 이미지화된 세계를 통해 드러나며 독자의 감각적 반응과 정서적 공명을 유도한다. 이는 ‘시화일여(詩畫一如)’로 대표되는 시인 정지용의 이미지즘적 창작 태도와도 맞닿아 있다. 정지용의 「비」에 등장하는 ‘수척한 흰 물살’과 같은 시어는 화자의 내면 정서를 감각적 이미지로 형상화한 사례라 할 수 있다(신용목, 2019). 이어서 신용목은 시의 이미지가 문자를 통해 생성된다는 점에서, 언어가 만들어내는 고유한 감각적·매개적 효과에 대한 새로운 해석이 필요하다고 보았다. 이에 본 연구는 이러한 문제

8) 김억은 시(詩)와 시가(詩歌)를 혼용하여 사용했다(구인모, 2002).

의식을 계승하여, 한국 근현대 시에 내재한 감정과 이미지를 언어 기호에 머무르는 텍스트 정보가 아닌, 디지털 방법론을 통해 색채라는 또 다른 감각 데이터로 전환하고자 한다.

(2) 문학 감정 데이터셋

본 절에서는 문학 텍스트의 감정을 다룬 데이터셋 사례를 살펴보고자 한다.

해외 연구에서는 문학 텍스트의 서사적 맥락과 감정 흐름을 반영하기 위해 다양한 데이터셋이 구축되어 왔다. 예를 들어 DENS(Dataset for Emotions of Narrative Sequences)는 영미권 소설을 대상으로 단일 감정 주석을 수행한 데이터셋이다(Liu et al., 2019). Picca와 Pavlopoulos(2024)는 고대 그리스 서사시 『일리아스』를 대상으로 다중 감정 라벨링을 적용한 데이터셋을 구축했다. 이와 더불어 시 텍스트를 대상으로 한 연구도 활발히 진행됐다. PO-EMO는 독일어 및 영어 시를 대상으로 행 단위의 다중 라벨링 방식을 도입한 감정 데이터셋이며(Haider et al., 2021), PERC는 인도 전통 미학에 기반을 둔 감정 체계를 적용하여 구축한 시 감정 데이터셋이다(Sreeja & Mahalakshmi, 2019). 이러한 선행 연구들은 문학이 가진 정서 효과를 온전하게 반영하기 위한 감정 주석 체계를 실험해 왔으며, 이는 문학 감정 데이터셋 구축에 있어 중요한 방법론적 기준을 제시한다.

한국의 경우, 일반어 기반의 감정 분석 연구가 이후 문학 텍스트 영역으로 확장되면서, 기존 감정 데이터셋을 문학 텍스트에 적용하려는 시도들이 나타났다. 초기에는 한국어 감정 어휘 사전과 같은 어휘 단위 자원이 주를 이루었으나(박상민 외, 2018; 손선주 외, 2012), 언어의 복합적인 맥락을 반영하는 데 한계를 지닌다는 점에서, 이를 보완하기 위해 영어 SNS 코멘트 기반 감정 라벨링 데이터셋인 GoEmotions(Demszky et al., 2020)을 번역·보정한 GoEmotions-Korean⁹⁾과 같은 데이터셋이 등장했다. 이후 연구는 단순 번역 수준을 넘어, 한국어의 문화적·정서적 특수성을 반영한 대규모 감정 데이터셋 구축으로 이어졌다. 대표적으로 KOTE(Korean Online That-gul Emotions) 데이터셋¹⁰⁾은 약 50,000개의 한국어 온라인 댓글을 대상으로 5명의 주석자가 44개 감정

9) GoEmotions-Korean 데이터셋은 다음 저장소에서 확인할 수 있다.
Park, J. (2021). *GoEmotions-Korean* [Data set]. GitHub.
<https://github.com/monologg/GoEmotions-Korean>

10) KOTE 데이터셋은 다음 저장소에서 확인할 수 있다.

범주에 따라 감정 라벨링을 수행함으로써 한국 사회의 다차원적인 감정을 포착했다 (Jeon et al., 2024). 또한, KOTE 데이터셋은 KcELECTRA¹¹⁾와 같은 언어 모델 파인 튜닝에 활용되어 보다 정교한 감정 분류 모델 개발에 도입됐다. 그러나 KOTE는 온라인 댓글을 중심으로 구성되어 있기 때문에, 은유와 상징 표현이 두드러지는 문학 텍스트의 정서적 특성을 충분히 반영하기에는 한계를 지닌다.

강우규(2024)는 KOTE의 44개 감정 범주를 참고하되, ‘애정’, ‘수치/죄책’ 등 문학적 맥락에서 두드러지는 감정 범주를 중심으로 총 18개의 감정 체계를 새롭게 설계하고(표 I-1 참고), 이를 바탕으로 고전 문학 『창선감의록』에 대한 감정 주석 데이터셋 KENT(Kang woo kyu Emotion for Novel Text)를 구축했다.¹²⁾

나아가 강우규와 김바로(2024)는 해당 분류체계와 데이터셋을 바탕으로 KcELECTRA 언어모델을 파인튜닝하여 개발한 감정 예측 모델의 성능을 검증했다. 이 연구는 기존 감정 분류 체계를 서사 중심 문학 텍스트에 적용하여 감정 분석을 진행한 연구 사례다.

[표 I-1] KOTE와 KENT의 감정분류체계 상관관계

KOTE	불평/ 불만	지긋지 긋	의심/ 불만	한심함	짜증	어이없 음	감동/감탄	고마움	존경
KENT	불만						감동/감사		
KOTE	환영/ 호의	안심/ 신뢰	슬픔	힘듦/ 지침	서러움	화남/ 분노	역겨움/ 징그러움	증오/ 혐오	경악
KENT	신뢰/호의		슬픔			분노/혐오			
KOTE	기대감	신기함/ 관심	아껴주는	흐뭇함(귀여움/ 예쁨)	기쁨	뿌듯함	편안/쾌적	즐거움/ 신남	행복
KENT	기대		애정		기쁨				

Jeon, D. (2024). *KOTE* [Data set]. GitHub. <https://github.com/searle-j/KOTE>

11) KcELECTRA는 한국어 온라인 뉴스 댓글 데이터를 기반으로 사전 학습된 한국어 언어모델이다. 자세한 내용은 다음 저장소에서 확인할 수 있다.
Lee, J. (2022). *KcELECTRA* [Software]. GitHub.
<https://github.com/Beomi/KcELECTRA>

12) KENT 데이터셋에 대한 자세한 내용은 다음 웹페이지에서 확인할 수 있다.
Kim, B., & Kang, W. (2025). *KENT (Kang woo kyu Emotion for Novel Text)* [Data set]. OSF. <https://doi.org/10.17605/OSF.IO/XSY8U>

KOTE	우쭐덤/ 무시함	안타까 움/실망	불쌍함/ 연민	부끄러움	절망	패배/자 기혐오	죄책감	공포/ 무서움	불안/ 걱정
KENT	우쭐덤/ 무시함	안타까 움/실망	불쌍함/ 연민	수치/죄책				두려움	
KOTE	없음	깨달음	당황/ 난처	부담/ 안_내킴	비장함	놀람	귀찮음	재미없 음	
KENT	없음	깨달음	당황/난처		비장함	놀람	-삭제-	-삭제-	

이처럼 문학 텍스트에 내포된 감정은 이분법이나 단일 감정으로 환원하기 어렵다. 특히 은유와 상징을 통해 다양한 정서가 중층적으로 교차하는 시의 경우, 이러한 한계는 더욱 두드러진다. 이에 본 연구는 필자가 참여한 선행 연구를 통해 구축한 한국 근현대 시 감정 데이터셋 KPoEM(Korean Poetry Emotion Mapping)과 감정 분류 모델을 연구 기반으로 삼아(Lim et al., 2026), 데이터셋 구축 과정과 활용 가능성을 검토하고, 이를 토대로 시 감정의 색채 시각화 환경을 구현하고자 한다.

(3) 문학 감정 데이터 시각화

문학 텍스트 감정을 시각화하려는 연구는 감정을 인지적 구조에서 감각적 경험으로 전환하고자 하는 흐름과 맞닿아 있다. 문학 감정 시각화 연구는 크게 세 가지 흐름으로 구분할 수 있다.

첫째는 감정의 시간적 흐름을 시각화하는 방식이다. 프로젝트 구텐베르크에 공개된 대규모 문학 코퍼스¹³⁾를 대상으로 감정 흐름을 분석한 Reagan 외(2016)의 연구는 1,300여 편의 소설에서 반복적으로 나타나는 감정 곡선(emotional arc)을 도출하고, 이를 상승형·하강형 등 몇 가지 패턴으로 분류했다. 이러한 연구는 감정이 서사 전반에 걸쳐 형성되는 동적 구조임을 보여주며, 이후 Öhman 외(2024)의 감정 곡선 연구는 9천여 편의 영어 소설을 대상으로, 감정의 분포와 변화 양상이 독자 수용 규모와 연관되어 있음을 실증적으로 제시했다. 이는 감정 데이터 시각화가 서사 구조

13) Project Gutenberg Genre Classification Corpus는 세계 최대의 자유 전자도서관인 Project Gutenberg(Hart, 1971)에 수록된 수천 권의 영어 고전 문학 작품을 대상으로 장르별 라벨을 주석한 데이터셋이다. 이 데이터셋은 주로 자연어처리(NLP) 분야에서 문학 텍스트의 장르 자동 분류 및 텍스트 스타일 분류 모델 학습에 널리 활용된다.
Hart, M. (1971). *Project Gutenberg*. <https://www.gutenberg.org/>

뿐만 아니라 문학적 경험과 해석의 층위까지 포착할 수 있음을 시사한다. 국내에서도 감정 분석 결과를 선형 그래프로 시각화하여 서사의 전개에 따른 감정의 상승과 하강을 추적하려는 연구가 수행됐다. 장세은 외(2021)는 소설과 영화 간 감정 구조를 비교하기 위해 이러한 선형 그래프 시각화를 활용했으며, 서사의 리듬과 패턴을 직관적으로 파악할 수 있도록 했다.

둘째는 감정을 등장인물 간의 구조적 관계로 모델링하여 시각화하는 방식이다. 박명건 외(2022)는 문장에 등장하는 감성 어휘의 점수를 단순 평균화하는 기존 감정 분석 방식의 한계를 보완하기 위해, 등장인물 간 감정 관계를 네트워크 구조로 모델링하는 방법을 제안했다. 이 연구에서는 개체명 인식을 통해 추출한 등장인물을 노드(node)로 설정하고, 동일 문장 내에서 함께 등장하는 인물들을 감성 정보를 반영한 엣지(edge)로 연결했다. 나아가 군집화 기법을 통해 인물 간 관계를 동조와 대립 그룹으로 분석함으로써, 서사 속 인물 관계의 변화 양상을 시각적으로 드러내고자 했다. 이러한 네트워크 기반 시각화는 문학의 감정을 개별 인물이나 문장의 속성으로 치환하지 않고, 인물 간 관계 요소로 해석한다는 점에서 의의를 지닌다. 특히 서사의 전개 과정에서 형성되는 갈등 구조와 집단 형성 양상을 직관적으로 파악할 수 있도록 하며, 감정이 서사 구조 속에서 어떻게 분포하고 변화하는지를 탐색할 수 있는 시각적 분석 결과를 제시한다.

셋째는 감정을 시각적 형태나 색채, 조형 요소로 변환하는 방식이다. 황정석과 조택연(2024)의 연구에서는 VAD(Valence, Arousal, Dominance)와 같은 다차원 감정 모델을 활용하여 텍스트의 감정을 정량화하고, 이를 조형 언어로 치환한 뒤, 인공지능을 활용해 시각적 형태를 생성하는 방법을 제안했다. 이 과정에서 감정은 언어와 시각 기호를 연결하는 매개로 작동하며, 이를 통해 텍스트의 분위기와 정서가 일관된 이미지로 구현될 수 있음을 확인했다. Wang 외(2025)는 문학 텍스트 기반 감정 분석 결과를 2차원 평면상에 색상을 연속적으로 매핑하는 프로그램 EmotionLens를 제안했다. 해당 연구는 Russell(1980)의 정서 원형 모델(circumplex model)¹⁴⁾을 기반으로 감정을 쾌-불쾌와 각성 정도를 두 축으로 구조화하고, 이를 색채 데이터와 연결함으로써 감정 분석 결과를 색채 표현으로 전환한다. 이러한 접근은 문학

14) Russell의 정서 원형 모델은 감정을 감정의 쾌-불쾌 정도(valence)를 x축에, 감정의 각성 정도(arousal)를 y축에 매핑하여 감정을 두 가지 연속적인 차원에 표현하는 모델이다.

텍스트의 감정을 정량적 색채 정보로 시각화했다는 점에서 의의를 지닌다. 그러나 감정을 동일한 명도(L=70)의 개별 색상 중심으로 환원하여 표현한다는 점에서, 명도 변화가 유도하는 감정 효과를 충분히 반영하지 못하며,¹⁵⁾ 서로 다른 정서와 심상이 중층적으로 교차하는 시 텍스트의 감정 구조를 포착하기에는 한계를 지닌다.

이에 본 연구는 시 텍스트에 내재한 복합적 감정과 심상을 드러낼 수 있는 배색 기반 감정 시각화 알고리즘을 설계하고자 한다. 특히 색의 명도와 채도 변화가 감정 반응에 영향을 미칠 수 있다는 선행 연구를 참고하여, 본 연구는 색조(명도·채도) 또한 감정 표현의 중요한 요소로 고려했다.

3) 시와 색채

(1) 시의 심상과 색채

시인은 감정, 사고, 상상 등을 전달하기 위해 시적 이미지를 활용하며, 이때 색채는 심상을 구성하는 중요한 매개로 기능한다. 색은 상징성과 정서적 함의를 통해 시의 분위기와 감정선을 형성한다. 한국 현대 시에서는 예를 들어, 푸른색은 ‘생명력’, ‘젊음’, ‘희망’, ‘이상’ 등을 암시하며, 붉은색은 ‘생명’, ‘사랑’, ‘열정’, ‘우국충절’을 연상시키는 등 특정 색채는 특정 감정이나 심리 및 상징과 결합해 의미를 구성한다(박종빈, 2010). 이문걸(2002)은 조지훈 시에 나타난 색채 감각을 분석하고, 색채가 시의 심미성과 심상을 이해하는 원형 이미지라고 보았다. 특히 조지훈 시에서 백색은 한국인의 전통적 미의식과 연결되며, ‘민족혼의 향수’와 ‘백의민족’, 그리고 민중의 정신성을 상징하는 이미지로 기능한다고 설명했다.

이러한 시의 색채 감각은 디지털 인문학 연구에서도 분석 대상으로 다루어져 왔다. 왕성필(2022)은 서정주의 시 텍스트에 나타난 색채 관련 어휘를 수집·분석하고 작품별로 색채 어휘의 출현 빈도를 계산했다. 이를 통해 해당 연구는 서정주 시 세계에서 나타나는 색채 감각의 변화를 시인의 시 세계 변천 과정으로 연결하여 해석했다. 이처럼 시의 색채 이미지는 시인이 구현하고자 하는 정서와 정신성, 문화적 정체성을 함축하는 감각적 기호로 기능한다.

15) 색의 명도와 채도와 같은 색채 속성의 변화는 사용자의 감정 반응에 영향을 미칠 수 있다(노연숙, 2015).

(2) 색채의 감성적 의미와 데이터화

색채는 인간의 생활과 감정에 밀접하게 연결된 지각적·직관적 기호다(여지혜 외, 2008). 인간은 대상을 바라볼 때 다른 요소보다 색상과 색조에 먼저 주목한다(한지운, 나건, 2021). 일반적으로 색채는 색상, 명도, 채도의 세 요소로 구성된다. 이 가운데 색상은 빛의 흡수·반사·투과 과정을 거쳐 인간의 눈에 지각되는 물체의 고유한 색의 성질을 의미하며, 명도는 색의 밝고 어두운 정도를, 채도는 색의 맑고 탁한 정도를 나타낸다. 특히 명도와 채도는 함께 색조, 즉 톤(tone)을 형성하며, 색채가 전달하는 분위기와 인상을 결정하는 중요한 요소로 작용한다. 색채는 물리적 파장의 차이를 넘어 인간의 문화적 경험과 심리적 반응에 따라 다양한 감성적 의미를 내포한다(Goethe, 2003). 이러한 색채와 감정의 상관성은 심리학, 미학, 디자인 등 다양한 분야에서 활발히 연구되어 왔으며, 최근에는 색채를 데이터로 구조화하고 인공지능 기술을 통해 분석·예측하려는 연구로까지 확장되고 있다.

정민영과 이현수(2019)는 배색 이미지에 감성 어휘를 대응시킨 라벨링 데이터셋을 구축하고, 이를 기반으로 배색 이미지에 대응하는 감성 어휘를 자동으로 예측하는 딥러닝 모델을 개발했다. 이 연구는 인간이 색채를 통해 지각하는 감성적 이미지를 인공지능 기술을 활용하여 ‘감성 어휘’라는 언어적 지표로 추출했다는 점에서 의의를 가진다. 나아가 색채의 감성 효과를 정량화하고 이를 기계가 처리 가능한 형태로 전환함으로써, 색채 심리 표현이 데이터로 해석될 가능성을 보여준다. 이는 색채가 심미성뿐 아니라 감각 및 감정을 전달하고 해석하는 매개로 기능하며, 문화적·심리적 차원이 복합적으로 작용한다는 점을 시사한다.

이러한 맥락에 비추어 볼 때, 본 연구가 시의 감정과 색채 이미지 간의 관계를 공감각적으로 시각화하기 위해서는, 인간의 해석이 반영된 정제된 데이터셋 구축이 필요하다. 따라서 본 연구는 기존 색채 심리 선행 연구에서 검증된 색채와 감정 대응 관계를 재구성하여, 한국인 색채 감정 데이터셋, KCoEM(Korean Color Emotion Mapping)을 구축하고자 한다.

(3) 배색(配色)과 감성·감정 형용사

색채는 단일 색의 상징적 의미뿐 아니라 배색을 통해 더욱 복합적인 감성을 유발할

수 있다. 김지혜와 박연선(2015)은 총 75개의 2색 배색 팔레트에 대해 한국인에게 인상 평가를 받은 결과, 2색 배색의 인상은 역량성, 화려함, 평가성 총 3가지 요소로 구성됨을 밝혔다(표 1-2 참고).

[표 1-2] 2색 배색 감성 평가 3요소

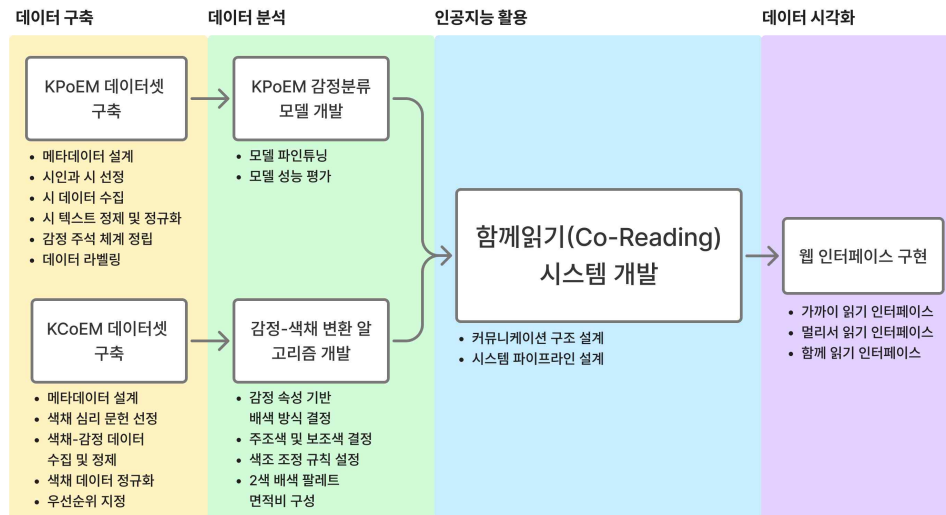
평가 3요소	의미	예시
역량성 (potency)	색을 통해 느껴지는 힘의 크기, 무게, 질감	‘부드럽다-딱딱하다’, ‘가볍다-무겁다’ 등
화려함 (colorful)	색을 통해 느껴지는 변화, 리듬과 같은 감각적 움직임	‘화려하다-수수하다’, ‘선명하다-흐릿하다’ 등
평가성 (evaluation)	조합된 배색이 조화로운지에 대한 가치 평가	‘조화-부조화’ ‘좋다-싫다’

이처럼 색채는 서로 조합될 때 단일 색상이 지니는 감정 효과를 증폭하거나 변환시키며, 명도와 채도의 조합에 따라 인상의 강도와 성격이 달라진다. 또한, 동일 색상 계열 배색은 대체로 조화로운 인상을 형성하는 반면, 대조 색상 계열 배색은 색조(톤)의 통일 여부에 따라 조화 또는 부조화의 인상이 달라질 수 있다. 이러한 결과는 배색 조화가 색채 인상에 크게 영향을 미칠 수 있음을 시사한다. 더불어 해당 연구는 한국어 문화권의 색채 감성 공간이 다른 문화권과 차이가 있음을 밝혀냈다. Ou 외(2004a, 2004b)는 단일 색상과 2색 배색이 유발하는 색채 감정을 각각 비교하고 실험했다. 이 연구는 색채 자극을 양극의 형용사 척도를 통해 체계적으로 측정할 수 있고, 색채 감정이 언어 표현과 밀접하게 연결되어 있음을 실증적으로 보여준다.¹⁶⁾

이처럼 색채는 특정 감정과 연상 작용을 유도하며, 형용사와 같은 언어 표현과도 밀접하게 연결되어 있다. 즉, 특정 색채나 배색 조합은 시각적 자극뿐만 아니라, 인간의 감정과 심리 상태를 표현하고 전달하는 감각적 매개다. 따라서 본 연구는 이러한 색채와 언어 간의 대응 관계를 기반으로, 한국 근현대 시의 감정을 색채 데이터로 변환하고, 인간과 인공지능이 언어 기반 해석 과정을 공유할 수 있는 배색 중심 데이터 시각화 시스템을 설계하고자 한다.

16) 이들의 배색 연구에서는 실험 참가자들이 총 190가지의 2색 배색 샘플에 대해 ‘따뜻한-시원한’, ‘무거운-가벼운’, ‘현대적인-고전적인’, ‘깨끗한-더러운’, ‘활동적인-수동적인’ 등 11쌍의 양극성 감정 척도를 평가했다(Ou et al., 2004b).

3. 연구 방법론



[그림 I-1] 연구 프로세스 개념도

본 연구의 목적은 시의 감정을 색채 이미지로 재구성하는 과정을 통해 데이터셋을 인간과 인공지능 사이의 해석과 상호작용을 매개하는 미디어 환경으로 전환하고, 문학 텍스트의 감정을 함께 읽고 탐색할 수 있는 커뮤니케이션 시스템, ‘함께 읽기 (Co-Reading)’을 제안하는 데 있다. 이를 위해 본 연구는 시-감정 데이터셋(KPoEM) 구축, KPoEM 감정 분류 모델 개발, 색채-감정 데이터셋(KCoEM) 구축, 감정-색채 변환 알고리즘 개발, 함께 읽기 시스템 설계, 웹 인터페이스 프로토타입 구현 과정을 통합적으로 수행한다(그림 I-1 참고).

① KPoEM 데이터셋 구축¹⁷⁾

먼저, KPoEM(Korean Poetry Emotion Mapping) 데이터셋 구축은 한국 근현대 시 텍스트를 대상으로 인간의 감정 해석을 데이터 형태로 구조화하는 과정이다. 본

17) KPoEM 데이터셋은 연구 활용 및 모델 배포를 위해 Hugging Face에 공개했으며, 데이터셋의 학술적 아카이빙과 인용, 연구 성과의 보존을 위해 Zenodo 저장소에도 등록했다. AKS-DHLAB. (2025a). *KPoEM* [Data set]. Hugging Face. <https://doi.org/10.57967/hf/6303>

Lim, I., Ji, H., Koo, S., Jung, S., Yun, J., & Kim, B. (2025). *KPoEM dataset* [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.15598092>

연구는 한국 근현대를 대표하는 시인 김소월, 윤동주, 이상, 임화, 한용운의 작품을 대상으로 수집한 시 텍스트를 정제한 뒤, 행 단위와 작품 단위 데이터로 구조화하고 감정 주석을 수행했다. 감정 주석은 KOTE 데이터셋의 44개 감정 범주를 기반으로 진행되었으며(Jeon et al., 2024), 총 5명의 주석자가 참여하여 동일 텍스트에 대한 최대 10개의 다중 감정 라벨링을 허용함으로써 문학 감정의 중층성과 해석의 다양성을 반영했다. 그 결과, 483편의 시를 대상으로 총 7,622건의 데이터를 구축했다.

나아가 본 데이터셋은 감정 분류 모델 학습의 기반으로 활용되며, 인간의 해석이 인공지능 시스템 안에서 다시 작동할 수 있는 해석 환경의 토대를 제공한다.

② KPoEM 감정 분류 모델 개발¹⁸⁾

다음으로, KPoEM 데이터셋을 활용하여 시 텍스트의 감정을 자동으로 분류할 수 있는 감정 분류 모델을 개발했다. 모델 학습에는 한국어 사전학습 언어모델인 KcELECTRA-base-v2022를 활용했으며, 데이터셋은 학습, 검증, 테스트 세트로 8:1:1 비율로 분할하여 학습을 수행했다. 또한, 최적의 하이퍼파라미터 조합을 탐색하기 위해 하이퍼파라미터 자동 최적화 프레임워크인 Optuna(v4.6.0)를 활용했으며(Akiba et al., 2019), KOTE만으로 학습한 모델, KPoEM만으로 학습한 모델, 그리고 KOTE와 KPoEM을 순차적으로 파인튜닝한 전이학습 모델의 성능을 비교·검증하였다. 모델의 성능은 Accuracy, Precision, Recall, F1-score, MCC(Matthews Correlation Coefficient)의 지표를 활용하여 평가했으며 실험 결과, KOTE 데이터로 사전 학습한 뒤 KPoEM 데이터로 추가 학습한 전이학습 모델이 가장 우수한 성능을 나타냈다. 이에 본 연구는 해당 모델을 최종 'KPoEM 감정 분류 모델'로 채택했다. KPoEM 데이터셋 구축과 감정 분류 모델 개발 과정은 본 연구자가 참여한 선행 연구의 성과를 기반으로 하며, 활용 코드는 GitHub 저장소에서 공개했다.¹⁹⁾

18) KPoEM 감정 분류 모델은 다음 저장소에서 확인할 수 있다.
AKS-DHLAB. (2025c). *KPoEM* [Software]. Hugging Face.
<https://doi.org/10.57967/hf/6301>

19) KPoEM 데이터셋 구축 및 감정 분류 모델 파인튜닝에 활용한 코드는 연구의 재현성과 활용을 위해 다음 저장소에 공개했다.
AKS-DHLAB. (2025b). *KPoEM* [Software]. GitHub.
<https://github.com/AKS-DHLAB/KPoEM>

해당 선행 연구는 국내외 학술대회 발표와 학술지 논문 게재 과정을 통해 검증 및 발전됐다(Lim et al., 2026).

본 연구에서는 KPoEM 감정 분류 모델을 활용하여 시 텍스트의 주 감정(primary emotion)과 보조 감정(secondary emotion)을 추출하고, 이를 감정-색채 변환 알고리즘의 입력값으로 활용했다. KPoEM 감정 분류 모델은 인간의 감정 해석을 학습한 인공지능 해석 장치이자 함께 읽기(Co-Reading) 시스템의 핵심 구성 요소로 기능한다.

③ KCoEM 데이터셋 구축²⁰⁾

이어서 한국인 색채 감정 데이터셋, KCoEM(Korean Color Emotion Mapping)을 구축했다. KCoEM은 KPoEM 데이터셋에서 사용된 KOTE의 44개 감정 범주를 기준으로, 국내외 색채 심리 관련 선행 연구를 종합하여 색채와 감정 간의 대응 관계를 체계화한 데이터셋이다.

데이터 구축 과정에서는 한국인을 대상으로 수행된 색채-감정 연상 실험 연구, 국내 색채 상징 및 감정 효과 관련 문헌, 그리고 기존 색채-감정 모델 등을 참고했으며, 서로 다른 감정 표현을 KOTE 감정 범주에 맞추어 정제 및 통합했다. 최종적으로 KCoEM은 총 348개의 색채 감정 데이터를 포함하며, 각 색채 정보는 RGB 값으로 정규화하여 전산적으로 활용할 수 있도록 했다. 또한 감정에 복수의 색채가 대응하는 경우를 고려하여 우선순위(priority) 메타데이터를 부여했으며, 한국산업표준(KS A 0011)에서 활용되는 먼셀(Munsell) 10색상환 체계를 참고하여 색상 간 관계를 구조화했다.

또한, KPoEM 데이터셋과 감정 분류 모델의 상세한 구축 과정, 그리고 RAG 기반 시 창작 모델의 활용 사례에 대해서는 다음 논문을 참고할 수 있다.

Lim, I., Ji, H., & Kim, B. (2026). KPoEM: A human-annotated dataset for emotion classification and RAG-based poetry generation in Korean modern poetry. *The Review of Korean Studies*, 29(1), 161-206.
<https://doi.org/10.25024/review.2026.29.1.006>

20) KCoEM 데이터셋은 다음 저장소에 공개하고 배포했다.

임이로. (2026). *KCoEM* [Data set]. Zenodo.
<https://doi.org/10.5281/zenodo.21131097>

이를 통해 KCoEM은 색채와 감정 대응 관계를 정량적으로 활용할 수 있는 데이터 기반을 제공하며, 본 연구의 감정-색채 변환 알고리즘의 핵심 데이터 자원으로 활용된다.

④ 감정-색채 변환 알고리즘 개발

시의 감정을 색채 이미지로 변환하기 위한 감정-색채 변환 알고리즘을 개발했다. 해당 알고리즘은 KPoEM 감정 분류 모델이 출력한 주 감정과 보조 감정을 입력값으로 받아, KCoEM 데이터셋을 기반으로 주조색과 보조색을 결정한다. 이 과정에서 KOTE의 감정 속성(긍정·부정·중립)과 색채 배색 원리인 유사색상배색 및 반대색상배색 규칙을 적용하여 감정 간 관계를 시각적으로 표현할 수 있도록 설계했다. 또한, 동일 감정에 복수의 색채가 대응하는 경우를 고려하여, KCoEM 데이터셋의 우선순위와 예외 처리 규칙을 적용해 주조색과 보조색 후보를 선정하도록 했다.

본 알고리즘은 Python(v3.10)²¹⁾ 환경에서 구현되었으며, 주조색과 보조색의 RGB 색상값을 HSB(Hue, Saturation, Brightness) 색공간으로 변환한 뒤 이미지 형용사를 기반으로 색조 조정 단계를 수행한다. 이를 통해 명도와 채도를 조정하여 감정의 강도와 분위기를 더욱 세밀하게 반영할 수 있도록 했다. 최종적으로는 색채의 조화도를 고려하여 주조색과 보조색 면적비를 6:4로 설정했다.

⑤ 함께 읽기(Co-Reading) 시스템 개발²²⁾

앞선 개발 단계에서 구축한 KPoEM 감정 분류 모델과 감정-색채 변환 알고리즘을 통합한 함께 읽기(Co-Reading) 시스템을 설계했다. 시스템은 사용자가 입력한 시 텍스트에 대해 감정 분류와 감정-색채 변환을 수행하고, 이를 바탕으로 최종 색채

21) Python Software Foundation. (2026). *Python* [Software].
<https://www.python.org/>

22) 감정-색채 변환 알고리즘과 함께 읽기(Co-Reading) 시스템 구현에 활용된 코드, 관련 참조 데이터 그리고 데이터 분석 코드는 다음 저장소에 공개했다.
임이로. (2026). *Co-Reading* [Software]. GitHub.
<https://github.com/poet-developer/Co-Reading>

팔레트를 생성하는 처리 구조를 가진다. 따라서 함께 읽기 시스템은 감정 분류, 감정-색채 변환, 컨텍스트 엔지니어링, 색조 조정, 색채 시각화 과정을 하나의 파이프라인으로 연결한다. 컨텍스트 엔지니어링에서는 Google Gemini-2.5-Flash LLM 모델을 활용하여 시 텍스트의 맥락적·감각적 특성을 반영하는 이미지 형용사를 선택하도록 설계했다. 이 단계에서 94개의 이미지 형용사와 12개의 형용사군, 그리고 141개의 배색 팔레트 데이터로 정제한 I.R.I 이미지 스케일(1998)을 외부 맥락 정보로 활용했다. 선택된 이미지 형용사는 형용사 군에 해당하는 배색 팔레트와 연결되며, 시스템은 이를 바탕으로 주조색과 보조색의 명도 및 채도를 조정한다.

이는 인간과 인공지능이 동일한 텍스트를 공유하며 각자의 해석 결과를 상호 참조할 수 있도록 커뮤니케이션 구조를 설계함으로써, 감정과 심상에 대한 새로운 해석 가능성을 탐색할 수 있는 ‘읽기’ 환경을 구현한 것이다. 한편, 본 연구에서 제안하는 KCoEM 데이터셋, 감정-색채 변환 알고리즘, 그리고 함께 읽기 시스템은 연구자가 관련 학술대회 발표를 통해 발전시켜 온 연구 결과물이다.²³⁾

⑥ 웹 인터페이스 구현

마지막으로 함께 읽기(Co-Reading) 시스템의 활용 가능성을 제시하기 위해 웹 기반 인터페이스 프로토타입을 구현했다. 본 인터페이스는 KPoEM 데이터셋을 대상으로 가까이 읽기, 멀리서 읽기, 그리고 함께 읽기의 세 가지 읽기 방식을 하나의 환경에서 시각적으로 제시하도록 설계했다.

인터페이스는 HTML, CSS, JavaScript를 활용한 정적(static) 프로토타입 형태로

23) 함께 읽기(Co-Reading)의 개념과 방법론, 그리고 기술 구현의 핵심 구조는 임이로(2025)의 학술대회 발표에서 처음 제안되었으며, 이후 국제학술대회에서 관련 내용을 소개했다 (Lim & Kim, 2026).

임이로. (2025). *데이터셋의 환경적 전환: 인공지능과 함께 읽기(Co-Reading)를 통한 한국 근현대 시-감정-색채 멀티모달 미디어* [Conference presentation]. 2025 중앙대학교 영화미디어학센터 대학원생 컨퍼런스. Zenodo.
<https://doi.org/10.5281/zenodo.17850832>

Lim, I., & Kim, B. (2026). *Co-Reading: A human-AI co-reader media system for poetry-emotion-color* [Poster presentation]. Hong Kong Association for Digital Humanities (HKADH), The Chinese University of Hong Kong (CUHK). Zenodo.
<https://doi.org/10.5281/zenodo.18753718>

구현했으며, GitHub Pages를 통해 배포했다.²⁴⁾

가까이 읽기 모듈은 KPoEM 데이터셋을 구성하는 다섯 명 시인의 작품에 대해 행 단위 감정 주석 데이터를 탐색할 수 있도록 일부 구현했으며, 멀리서 읽기 모듈은 감정 분포 그래프, 감정 히트맵, SHAP 분석 기반 워드클라우드를 통해 시인들의 전체 작품에 대하여 감정 구조를 거시적으로 탐색할 수 있도록 설계했다. 또한, 함께 읽기 모듈은 본 연구에서 생성한 시의 감정-색채 팔레트를 제시함으로써, 인간과 인공지능이 시를 함께 읽고 생성한 해석 결과를 비교·참조할 수 있도록 구성했다.

이로써 인터페이스는 데이터셋과 모델의 산출 결과, 그리고 사용자의 탐색 과정을 하나의 공간 안에서 조직한다. 따라서 데이터 시각화는 분석 결과를 제시하는 도구를 넘어, 다양한 해석 가능성을 탐색할 수 있는 읽기 환경으로 확장된다.

24) 본 연구에서 구현한 웹 인터페이스의 소스코드는 다음 저장소에서 확인할 수 있다. 또한 실제 웹 인터페이스 프로토타입은 GitHub Pages를 통해 공개 및 배포했다.

임이로. (2026). *KPoEMInterface* [Software]. GitHub.
<https://github.com/poet-developer/KPoEMInterface>

임이로. (2026). *KPoEMInterface* [Web app]. GitHub Pages.
<https://poet-developer.github.io/KPoEMInterface/>

II. 함께 읽기를 위한 데이터셋 및 모델 구축

1. 한국 근현대 시 감정 데이터셋 KPoEM 구축

1) KPoEM 데이터셋 개괄²⁵⁾

본 연구는 한국어 시 텍스트에 내재한 감정을 구조화하고, 이를 해석 가능한 데이터로 활용하기 위해 한국 근현대 시 감정 데이터셋 KPoEM(Korean Poetry Emotion Mapping, 이하 KPoEM)을 구축했다. KPoEM은 한국 근현대 시의 감정을 정량적으로 분석할 수 있는 기반을 제공하며, 이는 시 문학에 적합한 감정 분류 모델 개발로 이어진다.

KPoEM 데이터셋은 1920~1940년대를 대표하는 시인(김소월, 윤동주, 이상, 임화, 한용운)의 작품 483편을 대상으로 구축했다. 기존 KOTE 데이터셋(Jeon et al., 2024)의 설계를 참고하여 감정 주석 단위를 행 단위와 작품 단위로 구분했으며, 행 단위 데이터 7,007개와 작품 단위 데이터 615개를 포함한 총 7,622개 데이터로 구성했다.

행 단위 데이터셋은 시 본문을 행 단위로 분절한 뒤 무작위로 배열함으로써, 주석자가 전체 작품의 맥락에 영향을 받지 않고 개별 표현에 집중하여 감정을 주석할 수 있도록 설계했다. 반면 작품 단위 데이터셋은 시 전체의 문맥을 고려한 주석자의 감정 해석을 반영하기 위해 각 작품을 하나의 데이터 단위로 구성했으며, 위키문헌 원문의 행과 연 구조를 유지하여 최대한 형식과 리듬을 보존했다. 단, 작품 길이가 512자를 초과하는 경우 문단 단위로 분절했다. KPoEM 데이터셋 예시는 다음 표 II-1a와 표 II-1b와 같다.

25) KPoEM 데이터셋은 다음 저장소에서 확인할 수 있다.

AKS-DHLAB. (2025a). *KPoEM* [Data set]. Hugging Face.
<https://doi.org/10.57967/hf/6303>

Lim, I., Ji, H., Koo, S., Jung, S., Yun, J., & Kim, B. (2025). *KPoEM dataset* [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.15598092>

[표 II-1a] KPoEM 행 단위 데이터셋 예시

line_id	poe_m_id	text	sub_title	title	poet	annotator_01	annotator_02	annotator_03	annotator_04	annotator_05
664 5	446	구십리(九十里)가 반(半)이라네		성공	한용운	깨달음, 안심/신뢰, 기대감, 뿌듯함, 아껴주는, 환영/호의	힘듦/지침, 부담/안_내킴, 귀찮음, 불안/걱정, 지긋지긋, 기대감, 환영/호의	불평/불만, 안타까움/실망	감동/감탄, 기대감, 기쁨, 깨달음, 즐거움/신남, 행복, 환영/호의	깨달음
561 9	405	운명을 이야기할 수 있는 말을 가진 것이,		지상 의 시	임화	고마움, 안심/신뢰, 비장함, 서러움, 깨달음	고마움, 뿌듯함, 아껴주는, 안심/신뢰, 행복, 흐뭇함(귀여움/예 뻔)	깨달음, 비장함	감동/감탄, 비장함, 깨달음, 기대감, 뿌듯함	신기함/관심, 감동/감탄, 존경
692 3	479	내시시각각에 늘어서서한시 대나눌변인트 집으로나를위 협한다.	육친	위독	이상	불안/걱정, 의심/불신, 힘듦/지침, 공포/무서움, 부담/안_내킴, 귀찮음, 패배/자기혐오	공포/무서움, 귀찮음, 지긋지긋, 짜증, 패배/자기혐오, 화남/분노, 힘듦/지침, 부담/안_내킴, 서러움, 슬픔	불쌍함/연민, 불안/걱정, 불평/불만, 짜증	불안/걱정, 불평/불만, 의심/불신, 절망, 증오/혐오, 화남/분노	공포/무서움, 당황/난처

[표 II-1b] KPoEM 작품 단위 데이터셋 예시

seg_id	poe_m_id	text	sub_title	title	poetry_book	poet	annotator_01	annotator_02	annotator_03	annotator_04	annotator_05
285	234	의심하지 마세요 당신과 떨어져 있는 나에게 조금도 의심을 두지 마세요 의심을 둔대야 나에게는 별로 관계가 없으나 부질없이 당신에게 고통(苦痛)의 숫자(數)만 더할 뿐입니다 나는 당신의 첫 사랑의 팔에 안길 때에 온갖 거짓의 옷을 다 벗고 세상에 나온 그대로의 발가벗은 몸을 당신의 앞에 놓았습니다 지금까지도 당신의 앞에는 그때에 놓아둔 몸을 그대로 받들고 있습니다 만일 인위(人爲)가 있다면 「어찌하여야 처음 마음을 변치 않고 끝끝내 거짓 없는 몸을 나에게 바칠고」하는 마음뿐입니다 당신의 명령(命令)이라면 생명의 옷까지도 벗겠습니다 나에게 죄가 있다면 당신을 그리워하는 나의 「슬픔」입니다		의심하지 마세요	님의 침묵	한용운	비장함, 아껴주는, 환영/호의, 깨달음	서러움, 불평/불만, 불쌍함/연민, 슬픔, 안심/신뢰, 비장함, 슬픔, 존경, 죄책감, 아껴주는, 깨달음	감동/감탄, 불쌍함/연민, 비장함, 슬픔, 서러움, 안타까움/실망	슬픔, 비장함, 아껴주는	비장함, 슬픔, 아껴주는, 불평/불만, 서러움

		당신이 가실 때에 나의 입술에 수없이 입맞추고 「부디 나에게 대하여 슬퍼하지 말고 잘 있으라」고 한 당신의 간절한 부탁에 위반(違反)되는 까닭입니다								
286	234	그러나 그것만은 용서하여 주श्य 당신을 그리워하는 슬픔은 곧 나의 생명(生命)인 까닭입니다 만일 용서하지 아니하면 후일(後日)에 그에 대한 벌(罰)을 풍우(風雨)의 봄새벽의 낙화(落花)의 숫자만치라도 받겠습니다 당신의 사랑의 동아줄에 휘감기는 체형(體刑)도 사양치 않겠습니다 당신의 사랑의 혹법(酷法) 아래에 일만 가지로 복종(服從)하는 자유형(自由刑)도 받겠습니다 그러나 당신이 나에게 의심을 두시면 당신의 의심의 허물과 나의 슬픔의 죄를 맞비기고 말겠습니다 당신에게 떨어져 있는 나에게 의심을 두지 마श्य 부질없이 당신에게 고통(苦痛)의 숫자(數)를 더하지 마श्य	의심 하지 마श्य 요	님의 침묵	한용 운	아껴주는, 비장함, 슬픔, 서러움, 불안/걱정	당황/난처, 서러움, 불쌍함/연 민, 불안/걱정, 비장함, 존경, 뿌듯함, 슬픔, 아껴주는, 죄책감	감동/감탄, 불쌍함/연 민, 불안/걱정, 당황/난처, 비장함, 슬픔, 안타까움/ 실망	비장함, 슬픔, 안타까움/ 실망, 힘듦/지침	비장함, 깨달음, 불평/불만, 불안/걱정, 아껴주는, 안타까움/ 실망

KPoEM의 행 단위 데이터셋과 작품 단위 데이터셋은 각각 표 II-2a, 표 II-2b의 메타데이터 스키마에 따라 구성했다. 두 유형 데이터셋 모두 주석자가 본문을 우선적으로 읽을 수 있도록 식별자(ID) 다음에 본문(text)을 배치하고, 이후 제목(title)과 저자(poet) 정보를 배열했다. 또한, 이상의 「역단」과 같은 연작시의 경우, 소제목(sub_title) 항목을 추가하여 개별 시편을 구분했다.

[표 II-2a] KPoEM 행 단위 데이터셋 메타데이터 스키마

항목	설명
line_id	행 단위 데이터 고유 식별자
poem_id	데이터셋을 구성한 시 작품에 부여된 개별 식별자
text	시의 행
sub_title	연작시를 구성하는 개별 시편을 구분하기 위한 부제
title	시 제목 (예: 진달래꽃)
poet	시인 이름 (예: 김소월)
annotator_XX	해당 행에 대해 주석자가 부여한 감정 라벨 (예: annotator_01)

[표 II-2b] KPoEM 작품 단위 데이터셋 메타데이터 스키마

항목	설명
seg_id	작품 단위 데이터 고유 식별자 (512자 초과시 분절)
poem_id	데이터셋을 구성한 시 작품에 부여된 개별 식별자
text	작품 전체 본문 또는 512자 단위로 분절한 작품 데이터
sub_title	연작시를 구성하는 개별 시편을 구분하기 위한 부제
title	시 제목 (예: 서시)
poetry_book	시가 수록된 시집 또는 작품집 이름 (예: 하늘과 바람과 별과 시)
poet	시인 이름 (예: 윤동주)
annotator_XX	해당 작품에 대해 주석자가 부여한 감정 라벨 (예: annotator_05)

KPoEM 데이터셋 감정 주석에는 총 5명의 연구원이 참여했다. 주석자는 인문정보학 분야의 석·박사 과정생 및 졸업생으로 구성되었으며, 다양한 학문적 배경을 가지고 있다(표 II-3 참고). 또한 모두 한국어 모어 화자로서 문학 텍스트에 대한 독해 능력을 갖추고 있으며, 인문 데이터 연구 경험을 바탕으로 시 텍스트에 대한 감정 라벨링을 수행했다.

[표 II-3] KPoEM 데이터셋 주석자 구성

주석자	관련 전공 배경	학력	관련 활동 및 연구 경험
annotator_01	인문정보학	석사과정	문예 창작 및 인문 데이터 설계·구축 연구
annotator_02	인문정보학, 국어국문학	박사과정	한국 문학 및 전산 문학 연구
annotator_03	인문정보학	석사과정	인문 데이터 분석 연구
annotator_04	인문정보학	박사과정	인문 데이터 설계·구축 연구
annotator_05	인문정보학	석사졸업	인문 데이터 설계·구축 연구

주석자가 부여할 감정 라벨은 KOTE의 44개 감정 범주를 기반으로 구성했다(표 II-4 참고). 또한 KOTE 데이터셋과 동일하게, 각 행 또는 작품에 대해 최대 10개의 감정 라벨을 부여할 수 있는 다중 라벨 주석 방식을 채택했다.

[표 II-4] KOTE 44개 감정 어휘와 인덱스(id)

index	감정 어휘	index	감정 어휘	index	감정 어휘	index	감정 어휘
0	불평/불만	11	비장함	22	짜증	33	당황/난처
1	환영/호의	12	의심/불신	23	어이없음	34	경악
2	감동/감탄	13	뿌듯함	24	없음	35	부담/안-내 킴
3	지긋지긋	14	편안/쾌적	25	패배/자기혐 오	36	서러움
4	고마움	15	신기함/관심	26	귀찮음	37	재미없음
5	슬픔	16	아껴주는	27	힘듦/지침	38	불쌍함/연민
6	화남/분노	17	부끄러움	28	즐거움/신남	39	놀람
7	존경	18	공포/무서움	29	깨달음	40	행복
8	기대감	19	절망	30	죄책감	41	불안/걱정
9	우쭐땀/무시 함	20	한심함	31	증오/혐오	42	기쁨
10	안타까움/실 망	21	역겨움/징그 러움	32	흐뭇함(귀여 움/예쁨)	43	안심/신뢰

2) 데이터셋 구축 과정

KPoEM 데이터셋은 다음과 같은 순서로 구축 과정을 마쳤다. 1) 시인과 시 선정 기준을 결정하고 2) 위키문헌에서 시 데이터를 수집했으며, 3) 시 텍스트의 정제 및 정규화 과정을 거쳐 4) 감정 주석 체계에 따라 5) 5명의 주석자가 감정 라벨링을 수행했다.

(1) 시인과 시 선정

KPoEM 데이터셋 구축을 위해 시인 선정은 다음 세 가지 기준에 따라 결정했다. 첫째, 2000년부터 2019년 사이에 발표된 한국 현대문학 분야 박사학위 논문의 서지 데이터를 바탕으로 시인을 선정했다(김병준, 천정환, 2020). 특히 논문 제목이나 부제에 10회 이상 언급된 시인을 포함했으며, 이를 통해 학문적 대표성을 확보하고자 했다. 둘째, 한국시인협회가 발표한 ‘한국 시인 베스트 10’을 참고하여 대중적 인지도가 높은 시인을 추가로 선정했다(서울신문, 2007). 셋째, 데이터셋의 공개와 활용 가능성을 고려하여 사망 후 70년 이상이 지나 저작권 문제가 없는 대표적 한국 근대 시인을 선정했다. 이러한 기준을 바탕으로 김소월, 윤동주, 이상, 임화, 한용운을 대상 시인으로 선정했다.

시는 주로 위키문헌(Wikisource)에 공개된 시집에서 수집했다. 예를 들어, 한용운의 『님의 침묵』, 임화의 『현해탄』, 김소월의 『진달래꽃』, 윤동주의 『하늘과 바람과 별과 시』 등을 포함하며, 이상은 『위독』, 『오감도』, 『역단』 등 국문 연작시를 중심으로 선별했다. 최종적으로 김소월 165편, 윤동주 112편, 이상 46편, 임화 43편, 한용운 117편, 총 483편의 시를 수집했다.

(2) 시 데이터 수집

위키문헌은 위키미디어 재단이 운영하는 무료 온라인 디지털 도서관으로(그림 II-1 참고), XML 형식의 덤프 파일을 매달 두 차례 제공한다. 본 연구에서는 2024년 11월 1일 자 한국어 위키문헌(kowikisource) 덤프 파일을 활용하여 시 데이터를 수집했다.²⁶⁾ 덤프 파일은 Wikiextractor를 사용해 id, title, text 등 메타데이터를 포함

하는 JSON 형식으로 변환했으며, 표 II-5는 전처리 이전에 위키문헌에서 추출된 시
원문 텍스트 예시다.²⁷⁾



[그림 II-1] 위키문헌에 수록된 한용운, 「당신을 보았습니다」

[표 II-5] 한용운, 「당신을 보았습니다」 JSON 데이터 예시

<pre>{ "id": "3647", "revid": "5442", "url": "https://ko.wikisource.org/wiki?curid=3647", "title": "님의 침묵/당신을 보았습니다", "text": "&lt;poem&gt; 당신이 가신 뒤로 나는 당신을 잊을 수가 없습니다 까닭은 당신을 위하느니보다 나를 위함이 많습니다 나는 갈고심을 땅이 없으므로 秋收가 없습니다…&lt;/poem&gt;"} </pre>
--

26) 본 연구는 2024년 11월 1일자 한국어 위키문헌(kowikisource) 덤프 파일을 사용했다. 다만 위키미디어 덤프 파일은 운영 정책에 따라 갱신 또는 삭제될 수 있으므로, 추후 데이터 재현 시에는 위키미디어 덤프 저장소에서 제공하는 최신 버전의 파일을 사용하는 것을 권장한다.
Wikimedia Foundation. (2024). *Kowikisource dump files* [Data set].
<https://dumps.wikimedia.org/kowikisource/>

27) 한용운. (n.d.). *당신을 보았습니다*. 위키문헌.
https://ko.wikisource.org/wiki/님의_침묵/당신을_보았습니다

변환된 JSON 데이터는 pandas(v.2.3)²⁸⁾를 활용해서 DataFrame 형태로 변환했다. 이후, 본문(text) 데이터를 행 단위로 분절하고, id, text, title, poet 등 메타 데이터를 포함하는 기초 데이터 계층을 구축했다(표 II-6 참고).

[표 II-6] 한용운, 「당신을 보았습니다」 행 단위 분절 데이터 예시

lline_id	poem_id	text	sub_title	title	poet
2849	197	당신이 가신 뒤로 나는 당신을 잊을 수가 없습니다		당신을 보았습니다	한용운
2850	197	까닭은 당신을 위하느니보다 나를 위함이 많습니다		당신을 보았습니다	한용운
2851	197	나는 갈고심을 땅이 없으므로 秋收가 없습니다		당신을 보았습니다	한용운

(3) 시 텍스트 정제 및 정규화

수집한 시 텍스트는 라벨링 작업 이전에 다음과 같은 데이터 정제 및 정규화 과정을 거쳤다.

① 현대어 우선 선택과 한글(한자) 병기 규칙

시 작품에 전통 표기체 한국어와 현대 한국어 번역이 함께 수록된 경우, 데이터셋의 일관성을 유지하기 위해서 현대어 버전을 사용했다. 또한, 동일한 시의 여러 버전이 존재할 경우, 원문 표기법을 보존하고 한자 표기를 정확하게 포함한 버전을 우선으로 선택했다. 또한, 이 경우 한글-한자 병기 표현에 대해 일관된 규칙을 적용하여, 가독성과 원문 충실도를 모두 확보할 수 있도록 직접 정규화를 수행했다(표 II-7 참고). 또한 행 순서가 불일치할 경우, 수집한 데이터를 수작업으로 재정렬하여 원본의 순서와 일치시켰다.

28) pandas 라이브러리 관련 정보는 다음 웹 페이지에서 확인할 수 있다
The pandas development team. (2025). *pandas-dev/pandas: Pandas (v2.3)*. Zenodo. <https://doi.org/10.5281/zenodo.17229934>

[표 II-7] 한용운, 「타골의 시를 읽고」²⁹⁾ 한글-한자 병기 과정

전통 표기체 한국어	현대 한국어	연구 데이터 정규화 결과
벗이여 나의 벗이여 愛인의 무덤 위의 피어 있는 꽃처럼 나를 올리는 벗이여 작은 새의 자취도 없는 沙漠의 밤에 문득 만난 님처럼 나를 기쁘게 하는 벗이여 ...	벗이여, 나의 벗이여. 애인의 무덤 위에 피어 있는 꽃처럼 나를 올리는 벗이여. 작은 새의 자취도 없는 사막의 밤에 문득 만난 님처럼 나를 기쁘게 하는 벗이여. ...	벗이여 나의 벗이여 애인(愛人)의 무덤 위의 피어 있는 꽃처럼 나를 올리는 벗이여 작은 새의 자취도 없는 사막(沙漠)의 밤에 문득 만난 님처럼 나를 기쁘게 하는 벗이여 ...

② 데이터 제외

옛한글 형태(예: ㅎ)로 작성된 시는 현대 인코딩 환경에서 호환 문제로 인해 데이터셋에서 제외했다. 대표적으로 김소월의 「5일밤 산보」³⁰⁾가 있었으며 또한, 행 단위 데이터셋에서는 동일한 행이 반복되는 경우에는 그 중복을 제거했다. 한편, 이상의 「오감도 시제4호」와 「오감도 시제5호」와 같이 시각적 형식이나 이미지 요소를 포함한 작품은, 원문을 정확하게 재현하기 어렵다고 판단하여 데이터셋에서 제외했다.³¹⁾

③ 문자 정규화

앞선 한글(한자) 병기 규칙을 일괄 적용하되, 시 표현의 일관성을 확보하기 위해 추가적인 정규화 작업을 진행했다. 위키문헌에 이미 포함된 병기 표기(예: 단[熱한])는 그대로 보존했으며, 비문자 기호, 과도한 구두점(예: ……) 등 형식 요소들은 제거했다. 위키문헌 자체에 오타자가 발견된 경우, 해당 시집의 출판본을 참고하여 수정했다. 예를 들어, 잘못 표기된 ‘사디리’는 표준어 ‘사다리’로 정정했다.³²⁾

29) 한용운. (n.d.). *타골의 시를 읽고*. 위키문헌.

[https://ko.wikisource.org/wiki/님의_침묵/타골의_시\(GARDENISTO\)를_읽고](https://ko.wikisource.org/wiki/님의_침묵/타골의_시(GARDENISTO)를_읽고)

30) 김소월. (n.d.). *5일밤 산보*. 위키문헌. https://ko.wikisource.org/wiki/5일밤_산보

31) 이상. (n.d.). *오감도*. 위키문헌. <https://ko.wikisource.org/wiki/오감도>

32) 한용운. (2016). *님의 침묵 1*. 더스토리.

④ 시적 특성 보존

원문 시의 문학적 완성도를 유지하기 위해 시적 허용, 고어 표현, 그리고 고전적 어법은 현대어로 개작하지 않고 원형 그대로 보존했다. 예를 들어, ‘아니 눈물 흘리오리다’와 같은 표현은 원문의 정서와 운율을 유지하기 위해 수정 없이 유지했다. 또한, 한자어 반복을 나타내는 ‘々’와 같은 표기도 의미 왜곡을 방지하기 위해 그대로 보존했다.

⑤ 문장 분절 예외 처리

본 연구에 기본 분석 단위는 시의 ‘행’이다. 대부분의 시는 줄바꿈 문자(\n)를 기준으로 행 단위로 분절된다. 그러나 일부 시는 명확한 줄바꿈 없이 산문 형식으로 작성된 경우가 있었다. 이러한 경우, 텍스트가 네 문장을 초과하면 ‘-다’, ‘-요’, ‘-오리다’와 같은 문장 종결 어미를 기준으로 문장 단위로 분절했다. 구체적인 예시는 표 II-8a와 표 II-8b에 제시했다.

[표 II-8a] 이상, 「꽃나무」³³⁾ 원문 텍스트

대상 시 텍스트	제목	시인
별판한복판에 꽃나무하나가있소. 근처(近處)에는꽃나무가하나도없소. 꽃나무는제가생각하는꽃나무를 열심(熱心)으로생각하는것처럼 열심으로꽃을피워가지고셨소. 꽃나무는제가생각하는꽃나무에게갈수없소. 나는막달아났소. 한꽃나무를위(爲)하여 그러는것처럼 나는참그런이상스러운흥내를내었소.	꽃나무	이상

[표 II-8b] 이상, 「꽃나무」 문장 분절 예외 처리

line_id	poem_id	text	sub_title	title	poet
6699	450	별판한복판에 꽃나무하나가있소.		꽃나무	이상
6700	450	근처(近處)에는꽃나무가하나도없소.		꽃나무	이상
6701	450	꽃나무는제가생각하는꽃나무를 열심(熱心)으로생각하는것처럼 열심으로꽃을피워가지고셨소.		꽃나무	이상
6702	450	꽃나무는제가생각하는꽃나무에게갈수없소.		꽃나무	이상
6703	450	나는막달아났소.		꽃나무	이상
6704	450	한꽃나무를위(爲)하여 그러는것처럼 나는참그런이상스러운흥내를내었소.		꽃나무	이상

33) 이상. (n.d.). 꽃나무. 위키문헌. <https://ko.wikisource.org/wiki/꽃나무>

(4) 감정 주석 체계 정립

① 다중 라벨 주석 체계

한국 근현대 시의 미묘하고 복합적인 감정 표현을 포착하기 위해, 본 연구는 KOTE 데이터셋의 다중 라벨 주석 체계를 채택했다(Jeon et al., 2024). KOTE의 감정 분류 체계는 한국어 감정 어휘를 바탕으로 단어 임베딩 기법과 검증 절차를 거쳐 정제·군집화한 총 44개 감정 범주로 구성되어 있으며(표 II-4 참고), 기본 정서(예: 분노, 기쁨, 슬픔)는 물론 사회적·관계적 상태(예: 아껴주는, 환영/호의, 존경), 그리고 문화 특수적 감정(예: 비장함, 서러움 등)까지 폭넓은 정동 상태를 포괄한다. 본 연구에서는 이러한 44가지 감정을 시의 한 행 또는 한 작품에 대해 최대 10개의 감정을 동시에 부여할 수 있도록 했다.

② 시적 맥락 의존성 제어를 위한 무작위 배열

KOTE는 512자 이내의 댓글을 분석 단위로 삼아, 주변 문맥에 의존하지 않고 텍스트 자체에 내포한 감정을 분류하도록 구축된 데이터셋이다. 이와 유사하게 본 연구에서도 시의 각 행을 하나의 독립적인 분석 단위로 간주했다. 이를 위해 행 단위 데이터셋은 각 시의 행 순서를 무작위로 재배열한 셔플(shuffle) 형태로 구성했다. 이를 통해 주석자가 주변 구절이나 작품 전체의 맥락에 영향을 받지 않고 해당 행의 어휘적·문체적 특성만을 바탕으로 감정을 주석할 수 있도록 설계했다. 이는 시의 서사적 흐름과 구조적 연속성이 감정 해석에 미치는 영향을 통제하고, 감정 분류 모델 역시 행의 의미적 특성에 기반하여 감정을 학습하도록 하기 위함이다. 반면 작품 단위 데이터셋은 시 전체의 문맥을 고려한 감정 해석을 반영할 수 있도록 시 전체 맥락을 유지한 상태에서 감정 주석을 진행했다.

셔플한 행 단위 데이터셋은 데이터셋 배포 과정에서 line_id 순서로 재정렬했으며, 작품 단위 데이터셋은 배포용 버전에서 줄바꿈 기호(\n)를 제거하고 텍스트만 남긴 형태로 정리했다. 이는 데이터의 일관성과 활용성을 높이기 위한 조치로, 모델 학습 및 비교 실험 과정에서 줄바꿈 기호가 모델 학습 과정에 불필요한 영향을 미칠 수 있기 때문이다.

(5) 데이터 라벨링

5명의 주석자들은 각각 독립된 주석 환경에서 동일한 데이터셋에 대해 감정 라벨링을 수행했다. 주석 결과 간 의견 차이가 발생한 경우, 제3의 검토자가 교차 검토를 통해 이를 조정했다. 모든 주석자는 일관된 기준에 따라 주석을 수행할 수 있도록 사전에 표준화된 가이드라인을 바탕으로 교육을 받았다.

이 지침에는 1) 각 감정 라벨의 정의, 2) 시 데이터 주석 예시, 3) 행 단위 데이터셋을 먼저 주석한 뒤 작품 단위 데이터셋을 주석하도록 하는 작업 절차, 4) 데이터마다 최대 10개까지 감정을 복수로 평가하도록, 다중 라벨 주석 규칙을 포함했다.

3) 데이터셋 정량 분석 및 평가

이렇게 구축된 시인별 행 단위 데이터와 작품 단위 데이터의 통계는 다음 표 II-9와 같다. 전체적으로 데이터셋은 483편의 시에서 추출한 총 7,622개의 데이터로 구성되었으며, 감정 분포를 분석한 결과, 가장 빈도 높은 상위 10개 감정 범주는 표 II-10과 같다. 특히 불안/걱정, 슬픔, 기대감, 안타까움/실망과 같은 감정이 높은 빈도를 보이는데, 이는 식민지 경험, 전쟁, 근대적 감수성의 전환과 같은 한국 근대사의 역사적 맥락과 밀접히 연결된다고 해석할 수 있다.

[표 II-9] KPoEM 데이터셋 통계

개수	김소월	윤동주	이상	임화	한용운	통계	
행 단위	2,071	1,111	464	2,163	1,198	7,007	7,622
작품 단위	176	114	77	110	138	615	
작품수	165	112	46	43	117	483	

[표 II-10] KPoEM 데이터셋 빈도 상위 10개 감정 어휘

순위	감정 어휘	빈도
1	불안/걱정	10,126
2	슬픔	8,715
3	기대감	8,399

4	안타까움/실망	8,016
5	서러움	7,423
6	신기함/관심	7,094
7	비장함	6,808
8	아껴주는	6,786
9	감동/감탄	5,316
10	당황/난처	5,249

또한 주석자 간 감정 일치도를 분석한 결과, 전체 데이터의 약 99%에서 최소 두 명 이상 주석자가 동일한 감정에 합의한 것으로 나타났다(표 II-11 참고). 이는 데이터 셋이 주관적 감정 주석에도 불구하고 일정 수준 이상의 일관성을 확보하고 있음을 보여준다.

[표 II-11] 주석자 수에 따른 감정 주석 일치 통계

x명 이상 주석자가 동일하게 주석한 감정이 최소 하나 이상 존재하는 경우		x=1	x=2	x=3	x=4	x=5
데이터	개수(n)	7,622	7,613	7,052	4,659	1,725
	백분율(%)	100	99.88	92.52	61.13	22.63

2. KPoEM 감정 분류 모델³⁴⁾ 개발

본 연구는 KPoEM 데이터셋을 기반으로 한국 시 문학의 감정을 보다 효과적으로 반영하는 감정 분류 모델을 개발했다. 이를 위해 사전 학습된 트랜스포머(Transformer)³⁵⁾ 기반 언어 모델 KcELECTRA-base-v2022³⁶⁾을 파인튜닝³⁷⁾했다. 모델 학습 및 평가를 위해 KPoEM 데이터셋의 행 단위 데이터셋과 작품 단위 데이터셋을 각각 학습(train), 검증(validation), 테스트(test) 세트로 8:1:1 비율로 독립적으로 분할했다. 이후 동일한 분할 집합끼리 병합하여 최종 학습용 데이터셋을 구성했다(표 II-12 참고).

[표 II-12] KPoEM 데이터셋 학습·검증·테스트 세트 분할

항목	Train (80%)	Validation (10%)	Test (10%)	통계
개수	6,096	763	763	7,622

데이터셋의 감정 주석은 집계, 벡터화, 정규화 과정을 거쳐 모델 학습에 활용됐다. 먼저 각 데이터에 대해 5명의 주석자가 부여한 감정 라벨을 모두 집계한 뒤, 이를 44개 감정 범주에 대응하는 벡터로 변환하여 주석자 간 합의 정도를 모델 학습에 반영했다. 이후 Min-Max 정규화를 적용해 각 감정 수치가 0과 1 사이의 연속 값을 갖도록 조정함으로써 감정 간 상대적 중요도를 유지했다. 본 연구에서는 min-max 스케일링 계수로 KOTE 감정 분류 모델에서 사용한 0.2를 동일하게 사용했다. 이는 해당 계수가 KOTE 연구에서 경험적으로 도출된 값이며(Jeon et al., 2024), KOTE 연구와 동일한 실험 조건에서 성능을 비교·검증하기 위해 적용했다.

모델 학습 환경은 Python(v3.12.3) 기반으로 구축했으며, 하이퍼파라미터 최적화를

34) KPoEM 감정 분류 모델은 다음 저장소에서 확인할 수 있다.
AKS-DHLAB. (2025c). *KPoEM* [Software]. Hugging Face.
<https://doi.org/10.57967/hf/6301>

35) 트랜스포머(Transformer)는 입력된 텍스트의 각 요소 간 관계를 인간의 주의집중을 모방하는 어텐션(attention) 메커니즘으로 병렬적으로 학습하여 문맥을 이해하는 딥러닝 아키텍처를 의미한다(Vaswani et al., 2017).

36) KcELECTRA 모델은 다음 저장소에서 확인할 수 있다.
Lee, J. (2022). *KcELECTRA* [Software]. GitHub.
<https://github.com/Beomi/KcELECTRA>

37) 파인튜닝(fine-tuning)은 사전 학습된 언어 모델의 가중치를 새로운 데이터에 맞게 추가 학습하여 특정 작업에 특화시키는 전이학습 방식이다(위키백과, n.d.).

위해 Optuna(v4.6.0)³⁸⁾를 활용했다. 학습률(learning rate)은 $1e-6$ 에서 $5e-5$ 범위, 배치 크기(batch size)는 8에서 16 범위, 드롭아웃 비율(dropout rate)은 0.1에서 0.5 범위에서 탐색을 수행했으며, 각 실험은 3 epoch의 탐색 과정을 거쳐 최적의 하이퍼파라미터 조합을 도출했다. 모델 성능 평가는 763개 테스트 데이터를 사용하여 수행했으며, 분류 임계값(threshold)은 0.3으로 설정했다. 이는 전체 감정 범주 가운데 특정 감정 예측 점수가 0.3을 초과할 경우 해당 감정을 예측한 것으로 간주하는 기준을 뜻한다. 이 임계값 또한 KOTE 연구에서 제안된 값을 그대로 적용했다.

모델 학습은 세 가지 전략으로 적용하여 성능을 비교·검증했다. 표 II-13은 각 학습 전략에 따른 모델 성능 비교 결과를 제시한 것이다. KOTE 데이터셋만으로 학습한 모델은 상대적으로 낮은 성능을 보인 반면, KPoEM 데이터셋만으로 학습한 모델은 보다 안정적인 감정 분류 성능을 나타냈다. 특히 KOTE 데이터셋으로 사전 학습한 후 KPoEM 데이터셋을 추가 파인튜닝한 전이 학습 모델이 가장 높은 성능을 보였다.

[표 II-13] 감정 분류 모델 성능 평가 비교 (Threshold=0.3)

Model	Accuracy	Precision_micro	Precision_macro	Recall_micro	Recall_macro	F1_micro	F1_macro	MCC
KcELECTRA (KOTE)	0.77	0.49	0.46	0.38	0.33	0.43	0.34	0.29
KcELECTRA (KPoEM)	0.79	0.53	0.43	0.66	0.50	0.59	0.45	0.45
KcELECTRA (KOTE → KPoEM)	0.79	0.53	0.47	0.69	0.54	0.60	0.49	0.47

다음으로 각 실험 모델의 동작 특성을 정성적으로 평가하기 위해, 실제 한국 시 작품을 대상으로 감정 분류 결과를 비교했다. 분석 대상에는 2024년 노벨문학상 수상자인 한강(2013)의 「효에게. 2002. 겨울」과 한국 현대시 형성기를 대표하는 정지용의 「향수」³⁹⁾를 선정했다. 비교 분석 결과, 학습 데이터의 구성에 따라 모델 간 감정 예측 경향에 뚜렷한 차이가 나타났다(표 II-14 참고).

38) Optuna의 관련 정보는 다음 저장소에서 확인할 수 있다:
Optuna. (2025). *Optuna* (v4.6.0) [Software]. GitHub.
<https://github.com/optuna/optuna>

39) 정지용. (n.d.). 향수. 위키문헌. <https://ko.wikisource.org/wiki/향수/향수>

[표 II-14] 감정 분류 모델 정성 평가 비교

시 본문	제목	작가	KcELECTRA (KOTE)	KcELECTRA (KPoEM)	KcELECTRA (KOTE → KPoEM)
하지만 곧 너도 알게 되겠지 내가 할 수 있는 일은 기억하는 일뿐이란 걸 저 번쩍이는 거대한 흐름과 시간과 성장(成長), 집요하게 사라지고 새로 태어나는 것들 앞에 우리가 함께 있었다는 걸	효에게. 2002. 겨울	한강	깨달음: 0.80 기대감: 0.58 비장함: 0.55 안타까움/실망: 0.51 슬픔: 0.48 불안/걱정: 0.45 없음: 0.42 힘듦/지침: 0.39 절망: 0.32	비장함: 0.89 깨달음: 0.86 서러움: 0.77 불안/걱정: 0.72 안타까움/실망: 0.69 슬픔: 0.67 기대감: 0.63 힘듦/지침: 0.42 감동/감탄: 0.42 불평/불만: 0.33 의심/불신: 0.33 부담/안_내킴: 0.32	깨달음: 0.93 비장함: 0.91 기대감: 0.75 불안/걱정: 0.50 감동/감탄: 0.48 슬픔: 0.45 안심/신뢰: 0.44 안타까움/실망: 0.43 서러움: 0.39 기쁨: 0.39 환영/호의: 0.36 아껴주는: 0.35 뿌듯함: 0.32

흙에서 자란 내 마음 파아란 하늘 빛이 그립어 함부로 쓴 화살을 찾으려 풀섰 이슬에 함추름 휘적시든 곳, － 그 곳이 참하 꿈엔들 잊힐 리야. 전설바다에 춤추는 밤불결 같은 검은 귀밑머리 날리는 어린 누이와 아무렇지도 않고 여쁠 것도 없는 사철 발벗은 안해가 따가운 해스살을 등에 지고 이삭 줏던 곳, － 그 곳이 참하 꿈엔들 잊힐 리야.	향수	정지용	슬픔: 0.91 안타까움/실망: 0.72 불쌍함/연민: 0.66 절망: 0.56 힘듦/지침: 0.54 서러움: 0.49 불안/걱정: 0.42 깨달음: 0.35 없음: 0.31	안타까움/실망: 0.94 서러움: 0.94 슬픔: 0.94 불안/걱정: 0.85 아껴주는: 0.79 불쌍함/연민: 0.78 힘듦/지침: 0.68 깨달음: 0.63 기대감: 0.59 신기함/관심: 0.51 감동/감탄: 0.40 당황/난처: 0.38 부담/안_내김: 0.34 불평/불만: 0.33	슬픔: 0.97 서러움: 0.94 안타까움/실망: 0.90 불쌍함/연민: 0.75 불안/걱정: 0.68 아껴주는: 0.64 힘듦/지침: 0.63 기대감: 0.42 절망: 0.34 깨달음: 0.32 부담/안_내김: 0.31
--	----	-----	---	--	--

KOTE 데이터셋만으로 학습한 모델은 문맥상 부적절한 감정을 과도하게 예측하는 경향을 보인 반면, KPoEM 데이터셋만으로 학습한 모델은 시적 맥락을 보다 충실하게 반영한 감정 분포를 나타냈다. 특히 KOTE와 KPoEM을 차례로 학습시킨 전이 학습 모델이 가장 안정적이고 일관된 감정 예측 결과를 보였으며, 시의 미묘한 정서적 의미를 보다 정확하게 포착했다. 이러한 결과는 모델이 일반 도메인 감정 데이터를 먼저 학습하고, 문학 감정 데이터로 순차적 파인튜닝(sequential fine-tuning)을 수행하는 전략이 시 감정 분류에 효과적임을 보여준다. 본 연구에서는 이와 같이 최적화된 전이 학습 기반 모델을 'KPoEM 감정 분류 모델'로 정의한다.

이렇게 개발된 KPoEM 감정 분류 모델은 본 연구에서 개발한 감정-색채 변환 알고리즘에 활용된다. 즉, 모델이 입력한 시 텍스트에 대한 감정 예측 확률을 산출하고, 이 가운데 가장 높은 확률값을 보이는 감정을 주 감정(primary emotion), 두 번째로 높은 확률값을 보이는 감정을 보조 감정(secondary emotion)으로 정의하여 그 결과를 색채 시각화로 연결하는 핵심 기능을 수행한다. 본 모델은 KPoEM 데이터셋에 축적된 인간의 감정 해석을 학습한 결과물로서, 인간의 감정 해석을 인공지능이 재현하고 확장하는 해석 장치다.⁴⁰⁾

40) KPoEM 데이터셋 구축과 감정 분류 모델 파인튜닝에 활용한 코드는 다음 저장소에서 확인할 수 있다.

AKS-DHLAB. (2025b). *KPoEM* [Software]. GitHub.

<https://github.com/AKS-DHLAB/KPoEM>

KPoEM 데이터셋과 감정 분류 모델을 구축하고 개발하여 RAG 기반 시 창작 모델 개발에 활용한 사례에 대해서는 다음 논문을 참고할 수 있다.

Lim, I., Ji, H., & Kim, B. (2026). KPoEM: A human-annotated dataset for emotion classification and RAG-based poetry generation in Korean modern poetry. *The Review of Korean Studies*, 29(1), 161-206.

<https://doi.org/10.25024/review.2026.29.1.006>

3. 한국인 색채 감정 데이터셋 KCoEM 구축⁴¹⁾

1) KCoEM 데이터셋 개괄

본 연구에서는 감정을 색채로 시각화하기 위해, 한국인의 감정 범주와 색채 데이터를 연계한 한국인 색채 감정 데이터셋(Korean Color Emotion Mapping, 이하 KCoEM)을 구축했다. KCoEM은 KPoEM에서 사용한 KOTE 44가지 감정 범주(표 II-4 참고)에 대응하는 총 348개 색채 데이터로 구성했다. KCoEM은 색채와 감정 관계에 대한 인간의 문화적 경험과 해석이 축적된 지식을 구조화한 데이터셋이며, 기존 색채 연구에서 제시한 실험 결과와 해석을 전산적으로 재구성했다는 점에서 의의를 가진다.

데이터 구축에는 한국인을 대상으로 수행한 색채-감정 연상 실험 연구(성인 및 유아 대상), 국내 색채 상징 및 감정 효과 관련 문헌, 그리고 기존 색채-감정 모델 연구를 종합하여 구성했다. 수집한 감정 어휘는 각기 다른 선행 연구에서 사용한 표현과 범주 차이를 고려하여 KOTE 감정 클러스터 체계를 기준으로 재분류했으며(Jeon et al., 2024), 이를 통해 서로 다른 맥락에서 사용된 감정 표현을 일관된 44개 감정 범주와 대응시켰다. 직접적인 색채 대응 관계가 확인되지 않은 일부 감정에 대해서는 상위 감정군을 기준으로 색채 데이터를 보완했다.⁴²⁾ 색채에 대한 식별은 한국산업표준(KS A 0011⁴³⁾)에서 활용하는 먼셀(Munsell) 색체계⁴⁴⁾를 기준으로 삼았다. 또한, 서로 다른 색체계(KS, PCCS)의 색채 정보를 동일한 기준 안에서 비교하고 환

41) KCoEM 데이터셋은 다음 저장소에서 확인할 수 있다.

임이로. (2026). *KCoEM* [Data set]. Zenodo.
<https://doi.org/10.5281/zenodo.21131097>

42) 본 연구에서는 직접적인 색채 대응이 어려운 일부 감정에 대해, 최석재(2008)가 제시한 감정동사의 의미 유형 및 감정 분류를 참고했다.

43) KS A 0011은 한국산업표준(KS)에서 제정한 물체색의 이름 체계로, 먼셀(Munsell) 색체계를 바탕으로 색상을 분류하고 표기하기 위한 기준을 제시한다. 본 논문에서는 이를 ‘KS 색체계’로 약칭한다. 단, 본 연구에서는 기본 먼셀 10색상환(R, YR, Y, GY, G, BG, B, PB, P, RP)만을 사용하며, 분홍과 갈색은 제외했다.
국가기술표준원. (n.d.). *색채표준정보*. <https://www.kats.go.kr/content.do?cmsid=83>

44) 먼셀 색체계는 미국의 색채 연구가 먼셀(Albert H. Munsell)이 1905년 제안한 색체계로, 색상을 색상, 명도, 채도의 세 속성으로 체계화한 색채 체계이다. 색상은 10개의 기본 색상환을 기준으로 구성되며, 명도는 검정(N0)에서 흰색(N10)까지 단계적으로 구분된다. 채도는 색의 선명도를 의미하며 색상별로 서로 다른 범위를 가진다(이정아, 2013).

용할 수 있도록, 모든 색채 데이터는 RGB 값으로 변환 및 정규화하여 데이터셋으로 구조화했다.

KCoEM 데이터셋은 본 연구가 개발한 감정-색채 변환 알고리즘에서 감정 데이터를 색채로 변환할 때 기준 데이터로 활용한다. KCoEM 데이터셋 예시는 다음 표 II-15과 같다.

[표 II-15] KCoEM 데이터셋 예시

color_id	term_id	emotion_term	R	G	B	reference	source_system	reference_type	priority	color_label	color_code	KOTE_id
1	1	불만	101	49	121	정현희. (2018). <i>실제 적용 중심의 미술치료</i> (2판). 학지사.	국가기술표준원. (2009). <i>공공디자인 색채표준 가이드</i> . https://www.kats.go.kr/content.do?cmsid=85	국내 색채 상징 및 감정 효과 관련 문헌	1	보라	P	0
2	2	enthusiasm	185	32	33	Nijdam, N. A. (2006). <i>Mapping emotion to color</i> . University of Twente.	국가기술표준원. (2009). <i>공공디자인 색채표준 가이드</i> . https://www.kats.go.kr/content.do?cmsid=85	기존 색채-감정 모델	1	빨강	R	1
3	3	감동	190	26	52	김옥. (2017). <i>감정단어카드를 활용한 유아의 색채반응 연구</i> [석사학위논문, 광주여자대학교].	Wickham, C. (2024). <i>munsell: Utilities for using Munsell colours</i> (v0.5.1) [R package]. https://CRAN.R-project.org/package=munsell	한국 유아 대상 색채-감정 연상 실험	1	빨강	R	2

KCoEM은 선행 연구를 기반으로 색채-감정 대응 관계를 일관된 기준 아래에서 정렬하고 이를 재현 가능한 형태로 구조화한 데이터셋이다. 이러한 색채 데이터는 감정 어휘와 대응하는 색채 정보를 체계적으로 활용할 수 있도록 표 II-16에 제시한 메타데이터 스키마에 따라 구성했다.

[표 II-16] KCoEM 데이터셋 메타데이터 스키마

항목		설명
기본 식별 데이터	color_id	색채 데이터 고유 식별자
	term_id	감정 어휘 식별자
	emotion_term	색채에 대응하는 감정 어휘 (예: 기대하다, 분노, sadness)
색채 정량 데이터	R	RGB 색 공간 R(red) 좌표값
	G	RGB 색 공간 G(green) 좌표값
	B	RGB 색 공간 B(blue) 좌표값
문헌 기반 정보	reference	참조 문헌 인용 정보
	source_system	색채 정보 출처 및 변환 체계
	reference_type	참조 문헌 유형
연구자 주석	priority	동일 감정에 복수 색채 값이 존재할 경우 적용하는 우선순위
	color_label	계통색 이름 (예: 빨강)
	color_code	계통색 약호 (예: R)
	KOTE_id	대응하는 KOTE 감정 인덱스(ID) 번호

KCoEM 데이터셋은 각 색채 데이터와 감정 어휘를 연결하기 위한 기본 식별 정보를 포함한다. 또한, 문헌 기반 정보 항목들은 색채 데이터의 학문적 근거와 색채 정보 출처를 명시하기 위한 메타데이터다. 이러한 설계는 동일한 감정 어휘가 문화권이나 이론적 배경에 따라 서로 다른 색채 연상을 가질 수 있음을 고려한 것으로, KCoEM이 한국 문화권에 특화된 색채 감정 데이터셋으로 기능하는 중요한 기준이 된다. 특히, 색채 정보 출처 및 변환 체계(source_system) 항목에는 색채 값을 표준화하기 위해 활용한 색채 정보 출처 및 변환 체계 정보를 기록했으며, 참조 문헌에 RGB 값이 직접 제시된 경우에는 공란으로 처리했다.

연구자 주석 항목에서 계통색 이름(color_label)은 사람이 데이터를 직관적으로 이해할 수 있도록 돕는 설명 정보이며, 계통색 약호(color_code)는 프로그래밍 환경

에서 색상을 구분하기 위한 식별 정보다. 또한, KOTE_id는 KPoEM 감정 분류 모델의 출력 결과와 색채 데이터를 연결하기 위한 참조 정보로 사용한다.

특히, 우선순위(priority) 항목은 하나의 감정에 여러 색채 데이터가 대응될 수 있다는 색채 심리 연구 특성을 고려하여 설계한 메타데이터다. 이 항목은 본 연구의 감정-색채 변환 알고리즘에서 복수의 색채 데이터 후보 가운데 주조색과 보조색을 선택하는 기준으로 활용한다.

2) 데이터셋 구축 과정

KCoEM 데이터셋은 다음과 같은 순서로 구축 과정을 마쳤다. 1) 색채 심리 관련 선행 연구를 검토하여 데이터 수집 대상을 선정하고, 2) 각 연구에서 제시한 색채-감정 대응 관계를 수집·정리했으며 3) 수집한 색채 데이터에 KOTE 감정 어휘를 라벨링하고 4) 같은 감정에 대해 색채 데이터가 복수로 존재하는 경우 데이터의 신뢰도와 활용 가능성을 고려하여 우선순위를 부여했다.

(1) 색채 심리 문헌 선정

KCoEM 데이터셋은 KPoEM 감정 분류 모델 및 감정-색채 변환 알고리즘과 연동을 고려해 설계했다. 이를 위해 한국인 대상 색채-감정 연상 실험 연구(성인 및 유아 대상), 국내 색채 상징 및 감정 효과 관련 문헌, 그리고 기존 색채-감정 모델 연구 결과를 수집했으며, KOTE 감정 범주와 직접적으로 대응 가능한 색채-감정 관계만을 선별하여 데이터로 구성했다.

① 색채-감정 연상 실험 연구 : 성인 대상

한지운과 나건(2021)은 한국 성인 174명을 대상으로 87개 감정 어휘를 수집한 뒤, 이 중 18개 주요 감정 어휘를 도출했다(표 II-17 참고). 이들은 KS 색체계의 13가지 색조 중 실험에 적합한 밝은(light), 선명한(vivid), 진한(deep), 3가지 톤과 무채색(하양·회색·검정)을 포함하여 총 33가지 색채를 선정했으며, 이후 한국 성인 200명을 대상으로 각 색채가 연상시키는 감정을 조사했다. 각 색상은 국문 색명, 약호, 색상, 명도, 채도, RGB 값으로 구성된 표 형태로 제시했다.

[표 II-17] 한지운과 나건(2021)의 18개 감정 어휘

감정 어휘	감정 범주
걱정하다	불안걱정
감사하다	감동감사
고민하다	미분화괴로움
귀찮다	권태싫증
기대하다	바람
기쁘다	기쁨
긴장하다	미분화괴로움
답답하다	미분화괴로움
만족하다	기쁨
미안하다	수치죄책
부럽다	선망질투
사랑하다	애정
외롭다	슬픔우울
자랑스럽다	성취
재미있다	재미
짜증나다	미움
편안하다	평안
화나다	분노혐오

김애경과 오윤경(2016)은 기존 연구를 바탕으로 대표적인 긍정 및 부정 정서 어휘를 선정한 뒤(표 II-18 참고), 해당 정서 어휘를 한국 성인 143명에게 자극어로 제시하고, 120개 색상 팔레트 가운데 연상되는 색을 자유 응답 방식으로 조사했다. 실험에 사용한 색상 팔레트는 해당 연구에서 ‘J사 120색’이라고 명시하고 있을 뿐, 구체적으로 어떤 색체계를 사용하고 있는지는 언급하고 있지 않다. 그러나 사용하고 있는 색채 표기 방식을 살펴보면, 색조 표기에 KS 색체계와 달리 Bright와 Strong 톤을 사용하고 있으므로 이는 PCCS(Practical Color Coordination System) 색체계⁴⁵⁾

45) PCCS는 1964년 일본색채연구소(Japan Color Research Institute)에서 제안한 색체계로, 색상을 24개의 색상환 체계로 구분하고, 명도와 채도를 단계적으로 체계화하여 표현한다(김학성, 1997).

에서 파생된 팔레트를 사용하고 있음을 추정할 수 있다. 따라서 본 연구는 해당 연구 실험 결과를 PCCS 색체계를 기준으로 재정렬했으며, 선행 연구에서 유의미한 결과로 제시한 감정별 색채/색조 반응 비율(%)을 수집하여 KCoEM 데이터셋 구축에 활용했다.

[표 II-18] 김애경과 오윤경(2016)의 긍정·부정 어휘

긍정 어휘	부정 어휘
기쁨	경멸
만족	고통
애정	공포
열중함	괴로움
자존	근심
좋아함	분노
축복	슬픔
	애통
	원한
	전율
	죄책감
	질투

② 색채-감정 연상 실험 연구 : 유아 대상

김옥(2017)의 감정단어카드를 활용한 유아의 색채반응 연구에서는 만 4~5세 유아 394명을 대상으로 6개 기본 감정 범주에 속하는 20개 감정 단어를 제시하고(표 II-19 참고), 각 감정에 대해 유아가 연상하는 색채 반응을 조사했다. 실험에서는 KS 색체계의 33개 색채 자극을 활용하여 감정 단어 카드를 대상에게 제시한 뒤, 해당 감정과 어울린다고 생각하는 색을 선택하도록 했다. 해당 연구에서는 각 감정 단어에 대해 유아들의 색채 반응을 통계적으로 집계하여 색채-감정 대응 관계를 제시했으며, 본 연구에서는 선행 연구자가 통계적으로 유의미하다고 판단하여 제시한 감정별 색채 반응 결과(%)를 수집하여 KCoEM 데이터셋을 구축했다.

[표 II-19] 김옥(2017)의 기본 감정 및 세부 감정 범주

기본 감정	기쁨						슬픔			분노		두려움		사랑		미움				
세부 감정	감동	기대감	기쁨	신남	자신감	즐거움	고통	슬픔	외로움	절망	분노	화	공포	부끄러움	죄책감	사랑	애정	미움	증오	질투

③ 국내 색채 상징 및 감정 효과 관련 문헌

정현희(2018)는 유사한 환경을 공유하는 사람들은 색에 대해 공통된 연상을 형성하며, 이러한 연상이 축적되어 색의 상징성을 형성한다고 보았다. 또한, 해당 문헌에서는 다양한 선행 연구에서 제시한 색채 연상 및 상징을 종합하여 정리했다. 본 연구에서는 해당 문헌의 색채별 연상 및 상징 의미와 KOTE 감정 범주 간의 대응 가능성을 검토한 뒤 KCoEM 데이터셋 구축에 활용했다(표 II-20 참고).

[표 II-20] 정현희(2018)의 문헌 중 KCoEM 구축에 활용한 항목

색명	색의 연상 및 상징
빨간색	행복, 사랑, 분노
파란색	신뢰감, 놀라움, 절망
노란색	희망, 기대, 기쁨, 행복, 귀여움, 외로움, 사랑
녹색	위로, 희망, 게으름, 공포, 즐거움, 기쁨, 행운
주황색	희열, 사랑, 따뜻함, 거부당함
분홍색	귀여움, 사랑, 행복, 달콤함
보라색	불만, 공포, 분노
갈색	외로움
흰색	실패, 희망
검은색	절망, 두려움, 슬픔, 불안, 공포, 중압감, 후회

문재호와 신진우(2021)는 색채에 따른 감정 효과를 정리하고, 유채색의 감정 효과를 고명도·중채도·저명도와 같은 색조 특성에 따라 제시했다. 다만 해당 연구에서는 명확한 색조를 제시하지 않았기 때문에, 본 연구에서는 색조 구분이 적용되지 않은 무채색에 대한 색채-감정 관계만을 활용하여 KCoEM 데이터셋을 보완했다(표 II-21 참고).

[표 II-21] 문재호와 신진우(2021)의 색에 따른 감정 효과 자료 발췌

색채	고명도	중채도	저명도
빨강	행복/온화함/애정	질투/축제/반항/열정 /애정/수치	힘/답답함/무거움
주황	기쁨/명랑/애정/ 희망	웃음/축제/기쁨/반항 /질투	가을/풍요/칙칙함/엄격 /중후함
...			
검정	신비/정숙/슬픔/불안/공포/울울함/강함/고상함/비참/증오/원한/영혼		
회색	평범/소극적/차분/쓸쓸함/안정/우울함/따분함/낙담/ 역경(불행)/슬픔/고독/패배/고난		
흰색	청순/평화/단순함/마음/경건함/양심/미래/선/자유/영혼		

④ 기존 색채-감정 모델⁴⁶⁾ 연구

앞서 한국인을 대상으로 한 색채-감정 연상 실험과 국내 색채 상징 및 감정 효과 관련 문헌에서 확보되지 않은 KOTE 감정 어휘에 대해서는, 기존 색채-감정 이론 및 체계를 종합적으로 정리한 Nijdam(2006)의 연구를 보완 자료로 활용했다. 이 연구는 괴테(Goethe)의 색채 이론을 비롯하여 Claudia Cortes, Naz Kaya, Color Wheel Pro, Shirley Willett 등의 색채-감정 대응 모델을 비교·정리한 연구다. 본 연구에서는 해당 모델들이 제시한 감정 어휘 중 KCoEM 데이터셋 구축에 필요한 항목을 선별하여 사용했다(표 II-22 참고).

46) 본 연구에서는 기존 연구에서 제시된 색채와 감정 간 대응 관계를 설명하는 다양한 이론 및 체계를 포괄하여 '색채-감정 모델'로 통칭한다.

[표 II-22] KCoEM 구축에 활용한 Nijdam(2006) 연구의 감정 어휘

색채-감정 모델	감정 어휘
Goethe의 색채-감정 모델	pleasant, joy, irritating, happiness, comfort
Claudia Cortes의 색채-감정 모델	embarrassed, anger, love, joy, determination, fear, happiness, sick, faith, sadness, arrogant, sorrow
Naz Kaya의 색채-감정 모델	anger, happy, comfortable, tired, excited, no-emotion, disgust, annoyed, sick, lonely, bored, sad, fearful
Color Wheel Pro 색채-감정 모델	anger, joy, happiness, distrust, trust, frustration, sadness
Shirley Willett의 색채-감정 모델	enthusiasm, rage, anger, pride, shame, fear, satisfaction

데이터 대상 선정 과정에서 활용한 참조 문헌 정보는 데이터셋 reference 항목에 APA 7th 형식으로 출처를 명시했다. 또한, 자료의 성격에 따라 reference_type 항목에는 네 가지 유형(한국 성인 대상 색채-감정 연상 실험, 한국 유아 대상 색채-감정 연상 실험, 국내 색채 상징 및 감정 효과 관련 문헌, 기존 색채-감정 모델)으로 구분하여 입력했다. 이러한 분류는 색채-감정 대응 관계를 통합하는 과정에서 자료의 신뢰도와 문화적 맥락을 고려한 우선순위(priority)를 설정하는 기준으로 활용한다.

(2) 색채-감정 데이터 수집 및 정제

KCoEM 데이터셋의 감정 어휘(emotion_term)는 색채 관련 선행 연구마다 사용한 감정 어휘의 범위와 표기 방식에 차이를 보였다. 예를 들어, ‘기대하다’는 KOTE 감정 범주인 ‘기대감’과 대응될 수 있으며, ‘기쁨’, ‘슬픔’과 같이 KOTE와 동일한 어휘도 포함된다. 따라서 본 연구에서는 감정 어휘를 체계적으로 정제하고 KPoEM 및 KOTE 감정 체계(표 II-4 참고)와 정합성을 확보하기 위해 KOTE 연구가 제시한 감정 어휘 클러스터(cluster)를 활용했다. 이 클러스터는 44개 감정 범주와 유사한 감정 어휘를 군집화하고, 각 범주에 속하는 대표적인 영어 예시 어휘를 함께 제시한다. 본 연구에서는 이러한 클러스터와 영어 뜻풀이를 중간 기준으로 삼아, KCoEM의 감정 어휘를 KOTE 감정 범주와 대응시켰다.

구체적으로는, KCoEM 데이터셋의 감정 어휘를 KOTE의 감정 영문 표기와 대응시키기 위해, 국립국어원이 제공하는 한국어-영어 학습사전⁴⁷⁾의 뜻을 참고했다. 이후 해당 KCoEM 감정 어휘가 KOTE 감정 어휘 클러스터 예시 어휘와 의미적으로 부합하는 지를 검토하고, 이를 기준으로 각 색채 데이터에 대응하는 KOTE 감정 ID를 라벨링 했다.

한편, 일부 KOTE 감정 어휘는 기존 색채 연구에서 직접적인 색채-감정 대응 관계를 확보하지 못했다. 이러한 경우, 최석재(2008)가 제시한 감정군 분류를 보조 기준으로 활용했다. 예를 들어 ‘불쌍함/연민’은 ‘슬픔’ 감정군에 속한다는 점을 고려하여 기존에 확보한 ‘슬픔’ 색채 데이터를 대응시켰다. 또한 ‘경악’은 ‘놀람’ 감정군으로 분류되는 점을 참고하여, 기존의 ‘놀람’ 색채 데이터를 활용했다.

이러한 과정은 색채-감정 대응 관계를 일관된 범주 체계로 정렬하여 감정-색채 변환 알고리즘에서 감정 데이터의 재현성을 확보하는 기반이 된다. KOTE 감정 어휘 클러스터를 활용한 KCoEM 감정 어휘 대응 관계는 다음 [부록 1]에 제시했다.

(3) 색채 데이터 정규화

색채 이론에서는 다양한 색체계(color system)가 제안되어 왔다. KS 색체계는 세 가지 속성(색상·명도·채도)을 기준으로 색채를 체계적으로 분류한다. 본 연구에서 구축한 KCoEM 데이터셋은 KS 색체계가 사용한 먼셀 10색상환을 기준으로, 수집한 색채 정보를 기계 가독형 데이터에서 가장 널리 사용되는 sRGB 색공간으로 변환 및 정규화했다(Anderson et al., 1996).⁴⁸⁾ 이를 통해 KCoEM 데이터셋은 이후 감정 분류 결과를 색채로 변환하는 프로그래밍 과정에서 안정적으로 활용할 수 있다.

47) 관련 자료는 다음 웹페이지에서 확인할 수 있다.

국립국어원. (n.d.). *한국어-영어 학습사전 (Korean-English Learners' Dictionary)*.

<https://krdict.korean.go.kr/eng>

48) 본 연구에서는 디지털 환경에서 가장 널리 사용되는 RGB 색공간 표준인 sRGB를 사용하며, 용어의 일관성을 위해 이하에서는 편의상 ‘RGB’로 표기한다.

[표 II-23] KS 색체계 계통색 이름과 계통색 약호

구분	유채색										무채색		
계통색 이름	빨강	주황	노랑	연두	초록	청록	파랑	남색	보라	자주	하양	회색	검정
계통색 약호	R	YR	Y	GY	G	BG	B	PB	P	RP	N9.5	N5	N0.5

표 II-23은 KS 색체계를 토대로 국내 공공디자인 색채표준 가이드(2009)에서 사용하는 계통색 이름과 계통색 약호를 나타낸 것이다. 본 연구에서는 수집한 데이터에 대해 메타데이터의 color_label 항목에는 계통색 이름을, color_code 항목에는 계통색 약호를 주석했다. 이 과정에서 색채 및 디자인 분야의 전문 지식을 바탕으로 계통색을 분류했다. 한편 무채색은 명도만 존재하므로 Neutral의 약호인 'N'으로 표기하며, 명도 11단계 수치로 표현한다. 다만 KCoEM 데이터셋에서는 활용 편의를 고려하여 이를 White, Gray, Black으로 구분하여 표기했다.

RGB 색공간은 Red(R), Green(G), Blue(B)로 구성된 세 가지 기본색 성분 조합을 통해 색을 표현한다. 각 성분은 0부터 255 사이의 정수 값으로 나타내며, 세 가지 값의 조합에 따라 하나의 색채가 결정된다. 예를 들어, (255, 0, 0)은 순수한 빨간색을 의미한다. 이러한 RGB 색 표현 방식을 활용하여 색채 값을 메타데이터의 R, G, B 항목에 각각 분리하여 정규화했다. 이러한 정규화 과정은 서로 다른 문헌에서 제시한 색채 정보를 동일한 색체계 안에서 정렬하고, 색채-감정 데이터를 일정한 형식으로 관리하기 위한 것이다.

선행 문헌에서 제시한 색채 정보를 KCoEM 데이터셋에 입력하는 과정은 제시된 색채 정보의 구체성에 따라 크게 세 가지 유형으로 구분할 수 있다.

① 선행 문헌에서 구체적인 RGB 값을 명시한 경우

이 경우, 문헌에서 제시한 색채 정보가 이미 정량화되어 있으므로, 추가적인 변환 과정 없이 그대로 KCoEM 데이터셋에 반영했다. 대표적으로 한지운과 나건(2021)의 자료를 예시로 들 수 있다(표 II-24 참고).

[표 II-24] 한지운과 나건(2021)의 실험 결과 예시

감정 어휘	감정범주	%	색조	색상	표기	색상값	명도값	채도값	R	G	B
감사하다	감동감사	14	밝은	빨강	R/lt	5	7	10	253	143	136
감사하다	감동감사	13.5	밝은	자주	RP/lt	5	7	8	224	152	178
감사하다	감동감사	12.5	밝은	파랑	B/lt	5	7	8	84	200	216

② 선행 문헌이 특정 색체계에 기반하여 색상 약호와 명도·채도를 명시한 경우

특정 색체계에 기반하여 색상 약호와 명도·채도 값을 제시한 경우, 해당 색체계에 따라 색채를 RGB 값으로 변환하여 데이터셋에 반영했다. 국내 산업과 연구에서 주로 사용하는 색체계는 KS 색체계와 일본 색채연구소가 개발한 PCCS 색체계가 있다(표 II-25 참고).

[표 II-25] KS 색체계와 PCCS 색체계 비교

	KS 색체계	PCCS 색체계
기본 색상환	10색상	24색상
명도	11단계	17단계
색조(tone)	13개	12개

KS 색체계와 PCCS 색체계는 색상, 명도 및 색조를 표기하는 방식에서 차이를 보인다. 따라서 본 연구에서는 이러한 색체계 간 차이를 고려하여, 서로 다른 표기 방식으로 제시된 색채 정보를 동일한 기준에서 비교하고 활용할 수 있도록 RGB 전산값으로 통일했다. 구체적으로 PCCS 색체계를 기준으로 색상 약호와 명도·채도만 제시한 경우에는 공개된 색상 대응 자료⁴⁹⁾를 참고하여 해당 약호에 대응하는 RGB 값을

49) 본 연구에서는 PCCS 색체계의 색상 및 톤 위치를 보다 일관되게 대응시키기 위해 공개된 PCCS Tone Map 시각화 자료와 색상 코드 데이터를 함께 참조했다. 구체적으로는 PCCS Tone Map의 색상 배치 구조를 기준으로, 공개 저장소에서 제공하는 PCCS 색상 코드 데이터를 대응시킴으로써 각 색상의 상대적 위치와 톤 정보를 도출했다. 사용한 자료의 웹 주소는 다음과 같다. 다만, 본 연구에서 사용한 색상 정보는 PCCS의 공식 표준 데이터를 직접 제공하는 것이 아니라, 공개된 시각화 자료와 색상 코드 데이터를 바탕으로 연구 목적에 맞게 구성한 참조 데이터임을 밝힌다.

Hitokun. (n.d.). *PCCS tone map* [Web app]. GitHub Pages.

<https://hitokun-s.github.io/old/tool/pccs-tone-map.html>

Saito, H. (2013). *PCCS color code data* [Data set]. GitHub Gist.

<https://gist.github.com/hidex777/5268678>

입력했다. 또한 KS 색체계를 기반으로 색상 약호와 명·채도가 제시된 경우에는 R(v4.5.2)⁵⁰⁾ 환경에서 munsell(v0.5.1) 패키지⁵¹⁾를 활용하여 해당 색 약호를 RGB 값으로 변환했다(표 II-26, 표 II-27, 표 II-28 참고).

[표 II-26] 김애경과 오윤경(2016)의 실험 결과 예시 (PCCS 색체계)

감정 속성	감정 어휘	색상 선정율	색조 선정율	색상	색조
긍정	자존	18.9	39.2	PB	V
긍정	자존	18.9	11.9	PB	Dp
긍정	자존	18.9	11.2	PB	S
긍정	자존	16.1	39.2	P	V
긍정	자존	16.1	11.9	P	Dp
긍정	자존	16.1	11.2	P	S

[표 II-27] 김옥(2017)의 실험 결과 예시 (먼셀/KS 색체계)

기본 감정	세부 감정	계통색 이름	색채 표기	응답률
기쁨	감동	빨강	5R 4/14	6.9%
기쁨	감동	연한 보라	5P 9/6	6.6%
기쁨	감동	자주	5RP 4/12	6.1%

[표 II-28] Naz Kaya의 색채-감정 모델⁵²⁾ 예시 (먼셀 색체계)

색 이름	색채 표기	감정
Blue	10B 6/10	calm
Purple	5P 5/10	tired

50) R Core Team. (2025). *R: A language and environment for statistical computing* (v4.5.2) [Software]. R Foundation for Statistical Computing. <https://www.R-project.org/>

51) munsell은 먼셀 색체계의 색상 표기를 디지털 색상값으로 변환하고 활용할 수 있도록 지원하는 R 패키지다. 관련 정보는 다음 웹 페이지에서 확인할 수 있다.
Wickham, C. (2024). *munsell: Utilities for using Munsell colours* (v.0.5.1) [R package]. <https://CRAN.R-project.org/package=munsell>

52) 본 자료는 Nijdam(2006)의 연구에서 발췌하여 제시했다.

③ 선행 문헌에서 색상 이름만 제시한 경우

선행 문헌에서 색상 이름만 제시하고 명도나 채도에 대한 정보가 제공되지 않은 경우(표 II-29 참고), 색채 데이터의 신뢰도를 확보하기 위해 공공디자인 색채표준 가이드(2009)를 참조하여 해당 약호에 대응하는 대표 RGB 값을 입력했다.

[표 II-29] 선행 문헌에서 색상 이름만 제시한 경우 RGB 값 변환 예시

색상	감정	선행 문헌	공공디자인 색채표준 가이드	
보라색	불만	정현희. (2018).	보라	R: 101, G : 49, B: 121
Blue	sadness	Nijdam, N. A. (2006).	파랑	R: 0, G : 102, B: 162
회색	패배	문재호, 신진우. (2021).	회색	R: 121, G : 121, B: 121

(4) 우선순위 지정

색채 데이터 정규화 과정에서 하나의 감정에 대해 여러 색채 값이 복수로 제시되는 경우가 존재했다. 따라서 본 연구는 이후, 감정-색채 변환 알고리즘이 일정한 기준에서 색채 데이터를 결정할 수 있도록 우선순위(priority) 기준을 설정했다. 우선순위는 문헌의 유형과 실험 응답률을 기준으로 산정했으며, 1부터 차례로 주석했다. 이때 priority 값이 낮을수록 높은 우선순위를 갖는다.

① 문헌 유형에 따른 우선순위

KCoEM 데이터셋은 문헌 유형에 따라 다음과 같은 우선순위를 설정했다(표 II-30 참고).

[표 II-30] 문헌 유형별 우선순위

1순위	2순위	3순위	4순위
한국 성인 대상 색채-감정 연상 실험	한국 유아 대상 색채-감정 연상 실험	국내 색채 상징 및 상징 관련 문헌	기존 색채-감정 모델

한국인의 색채 연상 경향을 우선으로 반영하기 위해, 한국인을 대상으로 수행한 색채-감정 연상 실험 연구를 1차 자료로 활용했다. 이 가운데 성인 대상 연구를 기준

자료로 설정하고, 유아 대상 실험 연구를 그다음 우선순위로 배치했다. 다음으로 한국 사회에서 형성된 색채의 문화적 의미를 반영하기 위해 국내 색채 연상 및 상징 관련 문헌을 참고했으며, 특정 감정에 대응하는 색채 정보가 충분하지 않은 경우에는 기존 색채-감정 모델을 후순위 보완 자료로 활용했다.

② 실험 결과를 제시한 문헌의 응답률 반영

KCoEM 데이터셋 구축에 활용된 연구들은 실험 대상과 조사 방식이 서로 다르므로 각 연구의 응답률을 직접 비교하거나 통합하기에는 한계가 있다. 그러나 감정별 색채 연상 실험에서 나타난 상대적 응답률은, 감정에 대응하는 색상을 선택하기 위한 참고 기준으로 활용할 수 있다. 이에 본 연구에서는 동일한 문헌 유형 안에서 하나의 감정에 대해 여러 색채 값이 제시된 경우, 각 연구에서 보고된 응답률을 바탕으로 색채 값에 우선순위를 부여했다(표 II-31 참고). 즉, 특정 색채의 응답 비율이 높을수록 해당 감정과의 관련성이 상대적으로 높다고 보았다.

[표 II-31] KCoEM 데이터셋 우선순위 예시⁵³⁾

color_id	term_id	emotion_term	R	G	B	reference	source_system	reference_type	priority	color_label	color_code	KO TE_id
155	45	절망	0	0	0	김옥. (2017)	Wickham, C. (2024).	한국 유아 대상 색채-감정 연상 실험	1	검정	Black	19
156	45	절망	190	26	52	김옥. (2017)	Wickham, C. (2024).	한국 유아 대상 색채-감정 연상 실험	2	빨강	R	19
157	45	절망	118	52	54	김옥. (2017)	Wickham, C. (2024).	한국 유아 대상 색채-감정 연상 실험	3	빨강	R	19

53) 논문의 가독성을 위해 reference와 source_system 항목은 저자-연도 형식으로 축약하여 표기했으며, 전체 참고문헌 정보는 배포된 KCoEM 데이터셋의 메타데이터에 APA 7th 형식으로 수록했다.

158	45	절망	128	128	128	김옥. (2017)	Wickham, C. (2024).	한국 유아 대상 색채-감정 연상 실험	3	회색	Gray	19
159	45	절망	0	102	162	정현희. (2018).	국가기술표 준원. (2009).	국내 색채 상징 및 감정 효과 관련 문헌	4	파랑	B	19
160	45	절망	14	14	14	정현희. (2018).	국가기술표 준원. (2009).	국내 색채 상징 및 감정 효과 관련 문헌	4	검정	Black	19
161	46	frustration	101	49	121	Nijdam, N. A. (2006).	국가기술표 준원. (2009).	기존 색채-감정 모델	5	보라	P	19

3) 데이터셋 정량 분석 및 평가

KCoEM 데이터셋을 구성하는 색채 자료 출처별 비중을 살펴보면, 한국 성인 대상 색채-감정 연상 실험 자료가 전체의 50.86%로 가장 높은 비중을 차지했으며, 한국 유아 대상 실험 자료(20.40%), 국내 색채 상징 및 감정 효과 관련 문헌(16.38%), 그리고 기존 색채-감정 모델(12.36%) 순으로 나타났다(표 II-32 참고).

[표 II-32] KCoEM 데이터셋 출처 유형별 구성 통계

출처 유형	한국 성인 대상 색채-감정 연상 실험	한국 유아 대상 색채-감정 연상 실험	국내 색채 상징 및 감정 효과 관련 문헌	기존 색채-감정 모델	통계
데이터 수(n)	177	71	57	43	348
비율(%)	50.86	20.40	16.38	12.36	100

이와 같은 구성은 색채 데이터를 특정 단일 출처에 의존하지 않고, 실험 데이터와 문헌 기반 지식을 함께 반영함으로써 색채-감정 관계를 다층적으로 포착하고자 한 것이다.

KCoEM 데이터셋을 구성하는 색상 범주별 분포를 살펴보면(표 II-33 참고), 빨강(R)

이 가장 높은 비율을 차지했으며, 노랑(Y), 주황(YR), 보라(P), 자주(RP) 등이 비교적 높은 비중으로 나타났다. 반면, 청록(BG), 연두(GY), 하양(White)과 같은 색상은 상대적으로 낮은 비율을 보였다. 즉, 일부 색상에 데이터 구성이 집중되는 경향을 보였다. 이러한 분포는 색채와 감정 간의 일반적인 연상 경향을 일부 반영하는 동시에, 데이터가 수집된 연구 출처와 실험 조건, 그리고 한국 문화권의 색채 연상 특성이 복합적으로 작용한 결과로 해석할 수 있다.

[표 II-33] KCoEM 데이터셋 색상 범주별 분포 통계

순위	색상	색상 약호	데이터 수 (n)	비율 (%)
1	빨강	R	59	16.95
2	노랑	Y	38	10.92
3	주황	YR	34	9.77
4	보라	P	33	9.48
5	자주	RP	31	8.91
6	초록	G	26	7.47
6	파랑	B	26	7.47
8	검정	Black	23	6.61
9	남색	PB	22	6.32
9	회색	Gray	22	6.32
11	청록	BG	14	4.02
12	연두	GY	11	3.16
13	하양	White	9	2.59

본 연구에서 구축한 KCoEM 데이터셋은 감정-색채 변환 알고리즘에서 감정 데이터를 시각적으로 변환하는 핵심 매개로 활용한다.

4. 감정-색채 변환 알고리즘 개발

1) 감정 속성 기반 배색 방식 결정

앞선 단계에서 KPoEM 감정 분류 모델의 출력 결과 가운데 가장 높은 예측 확률을 보인 감정을 주 감정(primary emotion), 두 번째로 높은 감정을 보조 감정(secondary emotion)으로 정의했다. KOTE의 감정 체계는 긍정, 부정, 중립의 세 범주로 분류되며(Jeon et al., 2024), 본 연구에서는 이러한 감정 속성 분류를 배색 방식 결정의 기준으로 활용했다(표 II-34 참고). 따라서 감정-색채 변환 알고리즘은 주 감정과 보조 감정의 속성을 바탕으로 배색 방식을 결정하고, 이를 통해 감정 간의 관계를 색채 이미지로 시각화한다.

[표 II-34] KOTE가 제시한 감정의 긍정·부정·중립 분류

긍정	감동/감탄, 고마움, 기대감, 기쁨, 뿌듯함, 신기함/관심, 아껴주는, 안심/신뢰, 존경, 즐거움/신남, 편안/쾌적, 행복, 환영/호의, 흐뭇함(귀여움/예쁨)
부정	경악, 공포/무서움, 귀찮음, 당황/난처, 부끄러움, 부담/안 내킴, 불쌍함/연민, 불안/걱정, 불평/불만, 슬픔, 서러움, 안타까움/실망, 어이없음, 역겨움/징그러움, 의심/불신, 짜증, 재미없음, 절망, 죄책감, 증오/혐오, 지긋지긋, 패배/자기혐오, 한심함, 화남/분노, 힘들/지침
중립	깨달음, 놀람, 비장함, 우쭐덤/무시함

이러한 KOTE 감정 속성 분류는 주 감정과 보조 감정 간의 관계를 2색 배색 팔레트로 변환하기 위한 1차적 기준으로 사용했다. 본 연구는 감정 속성의 조합에 따라 배색원리 가운데 유사색상배색과 반대색상배색을 적용하며, 중립 감정이 포함된 경우에는 특정 배색 방식을 적용하지 않고 별도의 예외 처리 규칙을 적용했다. 감정 속성에 따른 배색 방식의 구체적인 기준은 표 II-35에 제시했다.

[표 II-35] KOTE 감정 속성에 따른 배색 방식

주 감정	긍정	부정	긍정	부정	금/부정	중립	중립
보조 감정	긍정	부정	부정	긍정	중립	금/부정	중립
배색 방식	유사색상배색		반대색상배색		예외처리		

주 감정과 보조 감정이 동일한 감정 속성(긍정-긍정 또는 부정-부정)에 속하는 경우, 정서적 일관성을 시각적으로 표현하기 위해 '유사색상배색'을 적용했다. 반면, 서로 다른 감정 속성(긍정-부정 또는 부정-긍정)에 속하면 감정 간 대비를 강조하기 위해 '반대색상배색'을 적용했다.⁵⁴⁾ 한편, 주 감정 또는 보조 감정이 중립 범주에 속하는 경우에는 특정 배색 방식을 우선 적용하지 않고, 이후 색상 결정 단계에서 별도의 예외 처리 규칙을 적용했다. 이때 감정 '없음'은 감정 속성 분류에 포함되지 않으므로 색상 결정 대상에서 제외했다.

2) 주조색 및 보조색 결정

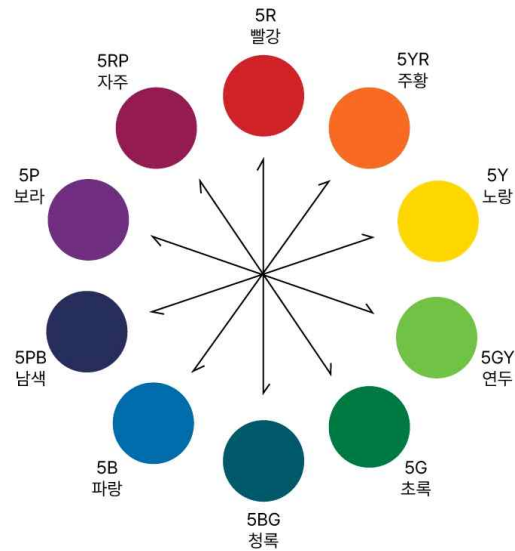
감정-색채 변환 알고리즘은 KCoEM 데이터셋을 검색하여 주 감정에 대응하는 색상 가운데 우선순위(priority)가 가장 높은 색상을 주조색으로 결정한다. 이후, 결정된 주조색을 기준으로 보조색을 결정하기 위한 후보 색상군을 탐색한다. 이러한 배색⁵⁵⁾ 기준은 먼셀(Munsell) 10색상환 체계를 기준으로 삼았다(그림 II-2 참고).

알고리즘은 앞 단계에서 결정된 배색 방식에 따라 주조색을 기준으로 보조색상 후보군을 구성한다. 유사색상배색이 적용된 경우에는 주조색과 인접한 유사색상군을 우선 탐색하며, 반대색상배색이 적용된 경우에는 주조색의 보색과 그 인접 색상으로 구성된 대조색상군을 우선 탐색한다.⁵⁶⁾ 예를 들어 주조색이 빨강(R)인 경우, 유사색상배색에서는 주황(YR)과 자주(RP)가 후보군에 포함되며, 반대색상배색에서는 보색인 청록(BG)과 대조색상군인 파랑(B), 초록(G)이 후보군에 포함된다.

54) 본 연구에서는 색채 이론의 유사색·보색·대조색 관계를 바탕으로, 감정-색채 변환 알고리즘의 구현을 위해 '유사색상배색'과 '반대색상배색'을 조작적으로 정의했다.

55) 배색에서 주조색(base color)은 가장 큰 면적을 차지하며 전체 색조의 분위기를 결정하는 중심 색을 의미한다. 보조색(dominant color)은 주조색 다음으로 높은 비율을 차지하는 색으로, 주조색과의 유사·대비·보색 관계를 통해 전체 배색의 조화를 형성하는 역할을 한다(위키백과, n.d.).

56) 유사색(analogous colors)이란 색상환에서 서로 인접한 색을 의미하며(위키백과, n.d.), 본 연구에서는 이러한 인접 색상들을 '유사색상군'이라 정의하고 배색에 활용한다. 또한, 보색(complementary color)은 주조색의 색상환의 정 반대에 위치한 색상을 말하며(위키백과, n.d.), 본 연구에서는 보색의 인접 색상들을 '대조색상군'이라고 정의하고 배색에 활용한다.



[그림 II-2] 먼셀(Munsell)의 10색상환

보조색은 앞서 구성한 보조색상 후보군과 보조 감정에 대응하는 KCoEM 색채 데이터를 교차 검색하여 보조색을 결정한다. 단, KCoEM 데이터셋에서는 하나의 감정에 복수의 색상이 대응될 수 있으므로, 보조색을 즉시 선택하지 않고 사전에 정의한 우선순위 규칙을 적용한다(표 II-36 참고). 이러한 규칙은 유사색상배색, 동일색상배색, 보색대비 및 대조색상배색과 같은 색상 대비 원리를 반영하여 설계했다(김미진, 2023; 김지혜, 박연선, 2015).

[표 II-36] 보조색 결정 우선순위

배색방식	보조색 후보군 우선순위		
	1	2	3
유사색상배색	유사색상군	동일색상	보조 감정에 대응하는 우선순위가 가장 높은 색채 데이터
반대색상배색	보색	대조색상군	
예외처리	보조 감정에 대응하는 우선순위가 가장 높은 색채 데이터		

유사색상배색의 경우, 유사색상군을 1순위로 적용하며, 해당 조건을 만족하는 색상이 존재하지 않을 경우 동일색상을 탐색한다. 이후에도 후보가 존재하지 않는 경우에는 주조색과 관계를 고려하지 않고, 보조 감정에 대응하는 색상 가운데 우선순위가 가장 높은 색채 데이터를 선택한다. 반대색상배색의 경우에는 보색을 1순위로 적용하고, 해당 색상이 존재하지 않을 경우 대조색상군을 탐색한다. 이후에도 적절한

후보가 존재하지 않는 경우에는 예외 처리 규칙에 따라 주조색과 관계를 고려하지 않고 보조 감정에 대응하는 색상 가운데 우선순위가 가장 높은 색채 데이터를 선택하도록 설계했다. 보조색 결정 단계에서 실험 결과의 재현 가능성을 확보하기 위해 난수 시드(random seed)를 42로 고정하여 실험을 수행했다.

3) 색상 결정 예외처리

주조색과 보조색 결정 단계에서는 KCoEM 데이터셋의 구조적 특성으로 인해 단일 규칙만으로는 색상 선택을 확정할 수 없는 예외적 상황이 발생할 수 있다. 이에 본 연구는 주조색과 보조색 결정 과정의 일관성과 안정성을 확보하기 위해 다음과 같은 예외 처리 규칙을 설계했다.

① 우선순위(priority)가 겹치는 경우

KCoEM 데이터셋은 복수의 선행 연구를 통합하여 구축했기 때문에 동일 감정에 대해 같은 우선순위를 갖는 색상이 존재할 수 있다. 이 경우 알고리즘은 동일 우선순위의 색채 데이터 가운데 하나를 무작위로 선택하도록 설계했다.

② 보조 감정에 대응하는 색상값이 복수인 경우

배색 방식에 따라 선택된 보조 색상 후보군에 복수의 색상이 존재하는 경우도 마찬가지로 우선순위를 기준으로 보조색을 결정한다. 이는 같은 감정에 대해 다양한 색채 반응이 존재할 수 있다는 점을 고려하면서도, 상대적으로 신뢰도가 높은 색채-감정 관계를 우선적으로 반영하기 위한 것이다.

③ 주 감정 또는 보조 감정이 중립 속성인 경우

주 감정 또는 보조 감정이 KOTE의 중립 감정 범주에 속하는 경우, 유사색상배색이나 반대색상배색과 같은 배색 원리를 우선 적용하지 않는다. 대신 해당 감정에 대응하는 색상 가운데 우선순위가 가장 높은 색상을 선택한다.

이와 같은 예외 처리 규칙은 주조색과 보조색 결정 과정에서 발생할 수 있는 불확정성을 제어하고, 다양한 조건에서도 안정적인 결과를 도출하도록 한다. 이는 연구

일관성과 재현 가능성을 확보하기 위해 규칙 기반 의사결정 과정을 명시적으로 설계한 것이다. 또한, 이는 감정-색채 변환이 KCoEM 데이터셋 구조와 배색 원리를 기반으로 우선순위 탐색과 예외 처리를 수행하는 규칙 기반 알고리즘임을 보여준다.

4) 색조 조정 규칙 설정

앞선 단계까지의 과정이 주조색과 보조색의 색상(hue)을 결정하는 과정이라면, 본 단계는 기존의 이미지 형용사와 배색 팔레트 데이터를 활용하여 명도(brightness)와 채도(saturation), 즉 주조색과 보조색의 색조를 조정하는 과정이다. 감정-색채 변환 알고리즘은 결정된 주조색과 보조색에 대해 인공지능 기반 컨텍스트 엔지니어링을 적용하여 색조 조정을 수행한다. 이때 인공지능은 I.R.I 이미지 스케일(1998)⁵⁷⁾의 이미지 형용사를 참조하여 입력된 시 텍스트를 읽고, 이에 대응하는 이미지 형용사를 한 개 선택한다. 선택된 이미지 형용사는 사전에 구조화된 배색 팔레트 데이터와 연결되어, 최종적으로 2색 배색 팔레트의 전체적 분위기와 색채 조화를 조정하는 기준이 된다.

색조 조정의 구체적인 과정은 제3장 함께 읽기(Co-Reading) 시스템 파이프라인 설계에서 예시와 함께 통합적으로 설명한다.

한편, 배색 팔레트 데이터는 RGB 값으로 정규화되어 있으나, 본 연구에서는 이를 파이썬 colorsys 라이브러리⁵⁸⁾를 활용하여 HSB 값으로 변환한다. HSB 색공간은 색상(Hue), 채도(Saturation), 명도(Brightness)를 독립적으로 표현할 수 있으므로, 색상(H)을 유지한 채 채도(S)와 명도(B)를 개별적으로 조정할 수 있다는 장점이 있다. 따라서 알고리즘은 주조색과 보조색을 HSB 색공간으로 변환한 후 색상(H)을 기준값으로 고정하고, 선택된 배색 팔레트의 채도(S)와 명도(B) 값을 참조하여 두 색상의 색조를 조정한다. 이를 통해 앞서 결정된 주조색과 보조색의 색상은 유지하면서

57) I.R.I 이미지 스케일은 한국인의 감성 구조를 기반으로 색채와 이미지의 관계를 체계화한 색채-감성 해석 체계로, 단색·형용사·배색 이미지 스케일로 구성된다.

58) colorsys 라이브러리는 Python에 내장된 표준 라이브러리로, 색상 체계 간 변환 기능을 제공한다. 관련 정보는 다음 공식 문서를 참고했다.
Python Software Foundation. (2026). *colorsys - Conversions between color systems*. In Python 3 documentation.
<https://docs.python.org/3/library/colorsyst.html>

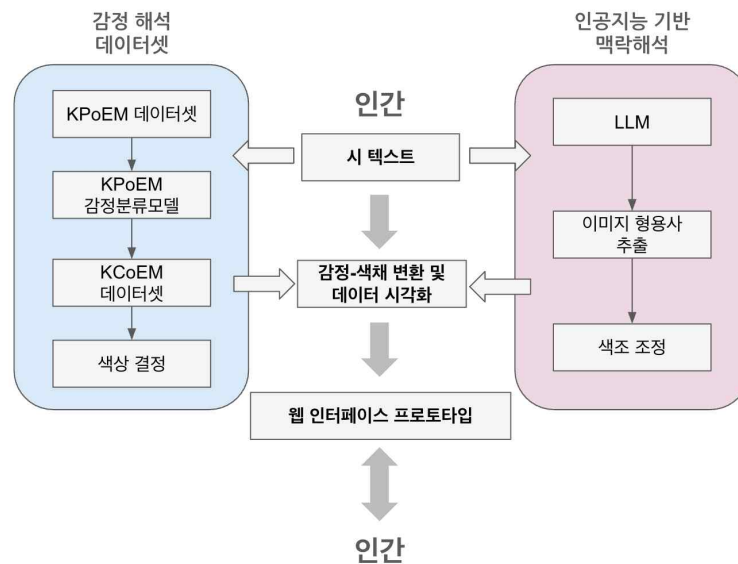
도, 인공지능이 해석한 이미지 형용사의 정서적 분위기를 반영한 색채 표현이 가능해진다. 조정된 주조색과 보조색의 HSB 값은 최종적으로 사용자가 확인할 수 있는 색채 팔레트 형태로 시각화하기 위해, 다시 RGB 색공간으로 변환한다.

5) 2색 배색 팔레트 면적비 구성

감정-색채 변환 알고리즘은 감정에 대응하는 주조색과 보조색을 먼저 선정한 뒤, 두 색상의 색조 조정 과정을 통해 배색의 조화도를 고려했다. 이는 배색의 조화도가 면적비 자체보다 어떤 색상이 선택되었는가에 더 큰 영향을 받기 때문이다(이윤진, 2016b). 따라서 본 연구는 최종적으로 출력된 2색 배색 팔레트가 일정 수준 이상 조화를 이룬다는 전제 아래, 선행 연구를 참고하여 2색 배색 면적비를 6:4(31~40%)로 설정했다(이윤진, 2016a). 이는 주조색이 시각적 중심 역할을 수행하면서도 보조색이 감정적 보완 기능을 수행하도록 고려한 것이다.

III. 함께 읽기(Co-Reading) 시스템 및 웹 인터페이스 구현⁵⁹⁾

1. 함께 읽기 시스템의 커뮤니케이션 구조 설계



[그림 III-1] 함께 읽기(Co-Reading)의 커뮤니케이션 구조

본 연구에서 제안하는 함께 읽기(Co-Reading) 시스템은 인간과 인공지능이 시를 매개로 상호작용하는 읽기 환경이다. 이 환경에서 시는 인간과 인공지능이 공유하는

59) 앞 절에서 설명한 감정-색채 변환 알고리즘 및 함께 읽기(Co-Reading) 시스템 구현을 위한 코드 전체는 다음 저장소에서 확인할 수 있다.
임이로. (2026). *Co-Reading* [Software]. GitHub.
<https://github.com/poet-developer/Co-Reading>

또한 본 연구는 임이로(2025)가 제안한 함께 읽기(Co-Reading)의 방법론과 시스템 설계를 확장·발전시킨 연구 결과물이다. 이후 국제학술대회에서 관련 내용을 소개했다 (Lim & Kim, 2026).

임이로. (2025). *데이터셋의 환경적 전환: 인공지능과 함께 읽기(Co-Reading)를 통한 한국 근현대 시-감정-색채 멀티모달 미디어* [Conference presentation]. 2025 중앙대학교 영화미디어학센터 대학원생 컨퍼런스. Zenodo.
<https://doi.org/10.5281/zenodo.17850832>

Lim, I., & Kim, B. (2026). *Co-Reading: A human-AI co-reader media system for poetry-emotion-color* [Poster presentation]. Hong Kong Association for Digital Humanities (HKADH), The Chinese University of Hong Kong (CUHK). Zenodo.
<https://doi.org/10.5281/zenodo.18753718>

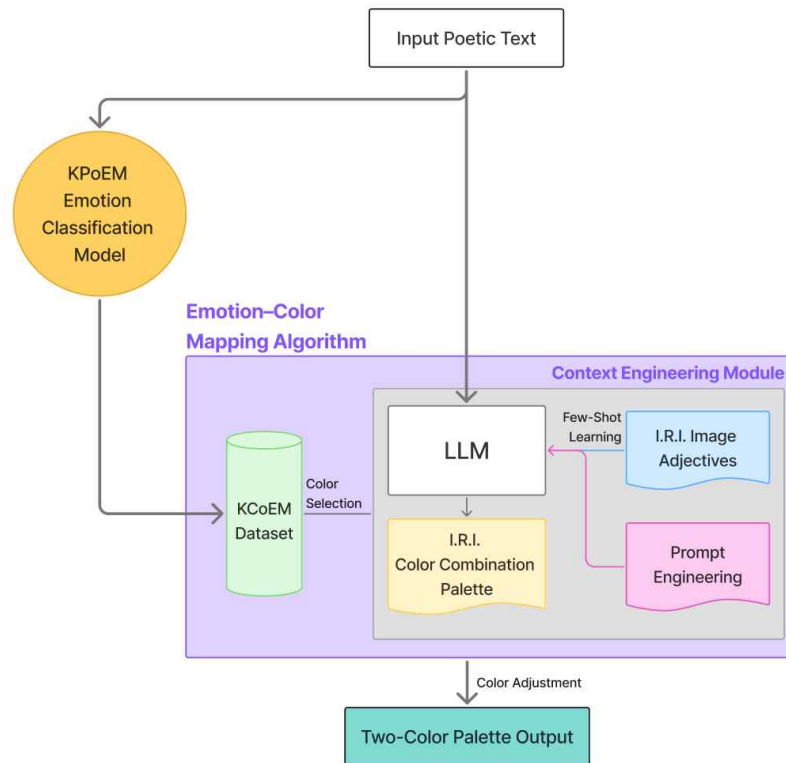
메시지로 기능하며, 시의 감정 정보는 색상(hue), 명도(brightness), 채도(saturation)로 구성된 색채 시각화 정보로 변환된다. 인간의 감정 해석이 축적된 KPoEM 감정 분류 모델과 KCoEM 데이터셋, 그리고 인공지능의 맥락적 해석은 이러한 변환 과정을 매개함으로써, 하나의 시를 함께 읽고 해석하는 다층적 상호작용 환경을 구성한다(그림 III-1 참고).

사용자가 입력한 시 텍스트는 먼저 KPoEM 감정 분류 모델을 통해 감정 정보로 출력된다. 본 연구에서 KPoEM 감정 분류 모델은 문학 텍스트에 대한 인간의 감정 해석을 학습한 모델로서, 시 텍스트의 정서적 특성을 전산적으로 재구성하는 해석 장치로 기능한다. 출력된 감정 정보는 KCoEM 데이터셋과 감정-색채 변환 알고리즘을 통해 색상 정보로 변환된다.

동시에 같은 시 텍스트는 대규모 언어 모델(LLM)에게 전달된다. LLM은 컨텍스트 엔지니어링(Context Engineering)을 통해 시 분위기와 심상을 반영하는 이미지 형용사를 선택하며, 해당 형용사는 감정-색채 변환 알고리즘에서 주조색과 보조색의 색조를 조정하는 기준으로 활용된다.

이처럼 함께 읽기(Co-Reading) 시스템은 시-감정-색채로 이어지는 변환 과정을 통해, 서로 다른 표현 체계 간의 의미를 번역하고 재구성한다. 이 과정에서 데이터셋은 단지 학습 자료나 참조 정보가 아니라, 인간과 인공지능이 시를 함께 해석하기 위해 구축한 환경적 미디어(environmental media)로 기능한다. 사용자가 시 텍스트를 시스템에 입력하면, 인공지능은 데이터셋과 알고리즘을 기반으로 이를 해석하고, 그 결과로 2색 배색 팔레트를 다시 사용자에게 반환한다. 이후, 사용자는 반환된 시각화 결과를 탐색하고 재해석함으로써, 의미 생성 과정에 다시 참여하게 된다. 따라서 본 연구에서 인공지능은 인간이 구축한 데이터셋과 해석 체계를 매개로 문학 해석에 참여하는 공동 독자(co-reader)로 기능한다.

2. 함께 읽기 시스템 파이프라인 설계



[그림 III-2] 함께 읽기(Co-Reading) 시스템 파이프라인 설계도

그림 III-2는 본 연구에서 구현한 함께 읽기(Co-Reading) 시스템의 전체 파이프라인을 나타낸 것이다. 사용자가 입력한 시 텍스트는 두 개의 경로로 처리된다. 첫째, KPoEM 감정 분류 모델을 통해 시 텍스트의 주 감정과 보조 감정을 추출하며, 그 결과는 KCoEM 데이터셋 기반의 감정-색채 변환 알고리즘에 입력된다. 둘째, LLM은 컨텍스트 엔지니어링을 통해 같은 시 텍스트의 맥락을 해석하고, I.R.I. 형용사 이미지 스케일의 이미지 형용사 가운데 하나를 선택한다. 이후, 감정-색채 변환 알고리즘은 선택된 이미지 형용사를 참조하여 최종 2색 배색 팔레트를 생성한다.

이러한 시스템의 작동 과정을 구체적으로 살펴보기 위해, 서정적 표현과 심상이 뚜렷하게 나타나는 한국 근현대 시인 신석정의 「네 눈망울에서는」⁶⁰⁾ 시 구절을 사례

로, 함께 읽기 시스템의 파이프라인을 단계적으로 설명하고자 한다. 이 과정은 1) 감정 분류 및 배색 방식 결정, 2) 주조색 및 보조색 결정, 3) 컨텍스트 엔지니어링 기반 이미지 형용사 선택, 4) 2색 배색 팔레트 선택, 5) 색조 조정 및 시각화 순서로 이루어진다.

1) 감정 분류 및 배색 방식 결정

함께 읽기(Co-Reading) 시스템의 첫 단계는 입력한 시 텍스트로부터 감정 정보를 추출하는 과정이다. 표 III-1은 신석정의 시 텍스트에 대해 KPoEM 감정 분류 모델이 분류한 상위 감정 결과를 제시한 것이다. 본 사례에서는 ‘흐뭇함(귀여움/예쁨)’이 0.97, ‘기쁨’이 0.89의 확률로 나타났으며, 그 외에도 ‘감동/감탄’, ‘신기함/관심’, ‘아껴주는’ 등의 감정이 높은 예측 확률을 보였다. 이 가운데 가장 높은 예측 확률을 보인 ‘흐뭇함(귀여움/예쁨)’과 두 번째로 높은 예측 확률을 보인 ‘기쁨’은 각각 주 감정과 보조 감정으로 결정된다.

또한, 주 감정과 보조 감정은 모두 감정 속성이 긍정 범주에 속하므로, 감정-색채 변환 알고리즘은 배색 방식을 ‘유사색상배색’으로 결정한다.

[표 III-1] 신석정, 「네 눈망울에서는」의 감정 분류 결과

사용자 입력 텍스트	KPoEM 모델이 분류한 감정 (상위 5개)	주 감정	보조 감정
네 눈망울에서는 초록빛 오월 하이얀 찔레꽃 내음새가 난다. 네 눈망울에는 초롱초롱한 별들의 이야기를 머금었다.	흐뭇함(귀여움/예쁨): 0.97 기쁨: 0.89 감동/감탄: 0.87 신기함/관심: 0.86 아껴주는: 0.82	흐뭇함(귀여움/예쁨)	기쁨
		주 감정 속성	보조 감정 속성
		긍정	긍정

60) 신석정 전집 간행위원회. (2009). *신석정 전집 1*. 국학자료원.

2) 주조색 및 보조색 결정

앞선 단계에서 신석정 시의 주 감정은 ‘흐뭇함(귀여움/예쁨)’으로 분류됐다. 시스템은 KCoEM 데이터셋에서 해당 감정과 대응하는 색상 후보를 검색한 뒤, 우선순위 규칙에 따라 주조색을 선택한다. 표 III-2는 흐뭇함(귀여움/예쁨)에 대응하는 KCoEM 데이터를 나타낸 것이다. 본 사례에서는 우선순위(priority)가 가장 높은 청록(BG)이 주조색으로 결정됐다.

[표 III-2] 신석정 사례의 주 감정에 대한 KCoEM 데이터

color_id	term_id	emotion_term	R	G	B	reference	source_system	reference_type	priority	color_label	color_code	KOTE_id
222	61	만족하다	0	151	133	한지운, 나건. (2021).		한국 성인 대상 색채-감정 연상 실험	1	청록	BG	32
223	61	만족하다	99	194	143	한지운, 나건. (2021).		한국 성인 대상 색채-감정 연상 실험	2	초록	G	32
...												
244	66	satisfaction	0	116	56	Nijdam, N. A. (2006)	국가기술표준원. (2009).	기존 색채-감정 모델	8	초록	G	32

표 III-3은 보조 감정인 ‘기쁨’에 대응하는 KCoEM 데이터를 나타낸 것이다. 앞선 단계에서 결정된 배색 방식은 ‘유사색상배색’으로, 본 사례에서는 ‘기쁨’에 대응하는 색상 가운데 주조색 청록(BG)의 유사색상군에 해당하는 초록(G)이 보조색으로 선택됐다. 비록 해당 색상 데이터의 우선순위는 10으로 상대적으로 낮지만, 감정-색채 변환 알고리즘은 KCoEM 데이터셋의 우선순위(priority)보다 사전에 결정된 배색 방식을 우선 고려하도록 설계했으므로, 초록(G)이 최종 보조색으로 선택된 것을 확인할 수 있다.

[표 III-3] 신석정 사례의 보조 감정에 대한 KCoEM 데이터

color_id	term_id	emotion_term	R	G	B	reference	source_system	reference_type	priority	color_label	color_code	KO TE_id
313	86	기쁨	238	201	0	김애경, 오윤경. (2016).	Hideo Saito. (2013).	한국 성인 대상 색채-감정 연상 실험	1	노랑	Y	42
314	86	기쁨	250	216	49	김애경, 오윤경. (2016).	Hideo Saito. (2013).	한국 성인 대상 색채-감정 연상 실험	2	노랑	Y	42
...												
323	86	기쁨	0	116	56	정현희. (2018).	국가기술표준원. (2009).	국내 색채 상징 및 감정 효과 관련 문헌	10	초록	G	42
345	79	희열	235	111	14	정현희. (2018).	국가기술표준원. (2009).	국내 색채 상징 및 감정 효과 관련 문헌	10	주황	YR	42
347	88	love	185	32	33	Nijdam, N. A. (2006)	국가기술표준원. (2009).	기존 색채-감정 모델	11	빨강	R	42

3) 컨텍스트 엔지니어링 기반 이미지 형용사 선택

최근 생성형 인공지능 연구에서는 프롬프트 설계와 외부 맥락 정보를 활용하여 모델의 성능을 향상시키는 컨텍스트 엔지니어링(Context Engineering)이 주목받고 있다(Mei et al., 2025). 본 연구에서는 LLM이 시 텍스트의 분위기과 심상을 해석하도록 하기 위해, 외부 맥락 정보로 국내 색채디자인 연구소 I.R.I가 개발한 형용사 이미지 스케일(1998)⁶¹⁾을 활용했다(그림 III-3 참고).

61) I.R.I 이미지 스케일은 한국인의 감성 구조를 바탕으로 구축된 색채-이미지 해석 체계로, 단색 이미지 스케일과 형용사 이미지 스케일, 그리고 배색 이미지 스케일로 구성된다. 본 연구에서는 그중 형용사 이미지 스케일과 배색 이미지 스케일을 활용했다. 이미지 스케일

부드러운
(soft)

가벼운 부드러운 투명한
아기자기한 상상한 밝은 상쾌한 맑은 시원한
즐거움 귀여운 사랑스러운 여성적인
재미있는 달콤한 향기로운 정돈된 깔끔한 순수한 온화한 약한
재활한 감미로운 감성적인 내추럴한 소박한 신선한
젊은 새로운 경쾌한 한상적인 포근한 중성적인 정중적인 관인한 인형된
스포티한 자유로운 장식적인 한국적인 여유있는 낙낙한
(動)동적인 울동적인 어우러는 넉넉한 우아한
선명한 다양한 태백적인 상쾌한 맛진
둔보이는 복잡한 화려한 시룩한 고급스러운 클래식한 은은한
귀여난 차가운 세련된 품위있는 조용한 수수한
개성적인 인공적인 하이테크한 미성숙한 격식있는 보수적인 고상한
다이나믹한 혁신적인 실용적인 친보적인 자적인 전통적인 자부한
거친한 강인한 서양적인 도시적인 점잖은 나이든 우물한
강한 거친 기능적인 견고한 중후한 딱한
외발드한 딱딱한 걸은 무거운 어두운

정적인
(Static)

IFU Design Institute Inc.
© 제 C-2001-001388호

I.R.I 형용사 이미지 스케일은 총 106개의 이미지 형용사를 12개의 이미지 형용사군으로 분류하고 있으며, 본 연구에서는 이를 인공지능의 해석을 위한 감성 어휘 체계로 활용한다. 감정-색채 변환 알고리즘은 컨텍스트 엔지니어링을 통해 LLM이 시를 읽고, I.R.I 이미지 형용사 가운데 하나를 선택하도록 설계했다. 선택된 이미지 형용사는 이후 알고리즘의 색조 조정 단계에서 명도와 채도 정보를 결정하는 기준이 된다. 즉, 이미지 형용사 데이터는 인공지능의 해석을 유도하는 맥락이다.

이를 통해 인공지능은 시 텍스트의 정서적 분위기와 심상을 해석하고, 그 결과를 색

디아리움. (2020). *IRI색채연구소 - 색채 이미지 공간*.
<https://www.engedi.kr/library/?bmode=view&idx=4236116>

조 조정 과정에 반영함으로써 동일한 감정 범주에 대해서도 서로 다른 시적 분위기를 색채로 표현할 수 있다.

따라서 본 연구는 I.R.I 이미지 형용사를 컨텍스트 엔지니어링에 활용하기 위해, 다음과 같은 과정을 거쳐 기계 가독형 데이터 JSON 파일로 구조화하고 정제했다. 해당 데이터는 함께 읽기(Co-Reading) 시스템 코드 저장소에 포함하여 공개했다.

(1) 이미지 형용사 데이터 구조화 및 정제

본 연구는 I.R.I 이미지 형용사 106개 가운데 일부 항목을 제외하여, 색채 매핑의 일관성과 해석적 타당성을 확보하고자 했다. 구체적으로, 정치적 의미를 내포하는 형용사(예: 보수적인, 진보적인), 성별·세대·문화적 고정관념을 강화할 수 있는 형용사(예: 남성적인, 여성적인, 나이든, 동양적인, 서양적인, 한국적인), 그리고 외래어 차용으로 인해 의미적 일관성이 저하될 수 있는 형용사(예: 와일드한, 심플한, 스포티한, 하이테크한) 등 총 12개 항목을 제외했다.

또한, 이미지 스케일 상에서 형용사가 포함되는 형용사군이 중복되거나 분류 경계가 불명확한 일부 항목(예: 맑은, 복잡한, 향기로운, 장식적인, 고급스러운, 중후한 등)은 연구자의 정성적 판단을 통해 하나의 군으로 재분류했다. 한편, ‘(활)동적인’ 형용사는 ‘활동적인’과 ‘동적인’으로 분리하여 모두 ‘경쾌한’ 군에 포함시켰다. 이와 같은 정제 과정을 거쳐, 최종적으로 이미지 형용사 94개를 형용사군 12개로 구성했다 (표 III-4 참고).

[표 III-4] 정제한 I.R.I 이미지 형용사

형용사군	이미지 형용사
귀여운	달콤한, 밝은, 사랑스러운, 싱싱한, 신선한, 아기자기한, 재미있는, 즐거운, 쾌활한
맑은	가벼운, 깔끔한, 맑은, 부드러운, 상쾌한, 섬세한, 얇은, 투명한
온화한	매끄러운, 소박한, 순수한, 안정된, 약한, 연약한, 유연한, 잔잔한
내추럴한	감미로운, 감성적인, 간편한, 자연적인, 전원적인, 정다운, 친근한, 편안한,

	포근한, 풍성한, 향기로운
경쾌한	다양한, 동적인, 돋보이는, 새로운, 선명한, 율동적인, 자유로운, 젊은, 활동적인
화려한	넉넉한, 매력적인, 복잡한, 성숙한, 시원한, 여유있는, 장식적인, 환상적인
은은한	가지런한, 그윽한, 단순한, 단정한, 정돈된, 정적인
우아한	감각적인, 고급스러운, 멋진, 편리한
다이나믹한	강한, 강인한, 개성적인, 거친, 기운찬, 뛰어난, 혁신적인
모던한	견고한, 기능적인, 깊은, 딱딱한, 도시적인, 무거운, 실용적인, 어두운, 인공적인, 차가운
점잖은	견실한, 격식있는, 세련된, 이성적인, 전통적인, 지적인, 품위있는
고상한	수수한, 오래된, 우울한, 조용한, 중후한, 차분한, 탁한

(2) LLM 프롬프트 설계

함께 읽기(Co-Reading) 시스템은 입력된 시 텍스트의 심상과 정서를 가장 잘 반영하는 I.R.I 이미지 형용사 1개를 선택할 수 있도록 설계됐다. 이때, 생성형 인공지능 모델은 Gemini-2.5-Flash⁶²⁾를 활용했으며, 94개의 이미지 형용사 가운데 하나만을 출력하도록 few-shot learning 기반 LLM 프롬프트를 설계했다(표 III-5 참고).

[표 III-5] 이미지 형용사 선택을 위한 LLM 프롬프트 구성

LLM 프롬프트
[역할] 당신은 시를 읽고 단 하나의 형용사 라벨을 부여하는 독자입니다. [지시 사항]

62) 본 연구에서 사용한 Gemini-2.5-Flash API에 대한 정보는 다음 웹 페이지에서 확인할 수 있다.

Google. (2026). *Gemini-2.5-Flash* [Large language model]. Gemini API.
<https://ai.google.dev/gemini-api/docs/models/gemini-2.5-flash>

1. 아래 제공된 '후보 라벨' 목록에서 '시'의 심상과 정서에 가장 부합하는 하나를 선택하세요.
2. ****반드시**** 목록에 있는 라벨만 사용하세요.
3. 결과물에는 선택한 ****라벨** 외에 어떤 설명이나 문장도 포함하지 마세요.**
4. 라벨의 형태(어미)를 바꾸지 말고 목록에 있는 그대로 출력하세요.

[후보 라벨]

{label_list_str}

[시]

{poem}

[출력 형식]

라벨 하나만 출력하십시오.

프롬프트 설계의 핵심은 인공지능의 역할을 ‘공동 독자(co-reader)’로 재위치 시키는 데 있다. 모델은 시 텍스트 전체를 맥락으로 입력받되, 출력 단계에서는 감정이나 이미지를 직접 생성하지 않고, 인간이 사전에 구축한 이미지 형용사 집합 가운데 하나를 선택하도록 제한함으로써, 본 시스템이 하나의 해석 환경으로 작동하도록 설계했다. 이러한 제약은 생성형 인공지능의 환각(hallucination) 현상을 억제하는 것이 아니라, 인간이 구축한 해석 환경 안에서 제한적으로 허용하기 위한 목적을 지닌다.

한편, 본 연구에서는 연구 재현성을 확보하기 위해 생성형 인공지능의 temperature 값을 0으로 설정하여, 동일한 입력에 대해 일관된 이미지 형용사 선택 결과를 도출할 수 있도록 했다.

(3) 신석정, 「네 눈망울에서는」 적용 결과

신석정의 「네 눈망울에서는」 시 구절을 프롬프트에 입력한 결과, LLM은 시 전반의 정서적 분위기와 심상을 반영하는 이미지 형용사로 ‘순수한’을 선택했다(표 III-6 참고).

[표 III-6] 신석정 사례의 감정 분류와 이미지 형용사 선택 결과

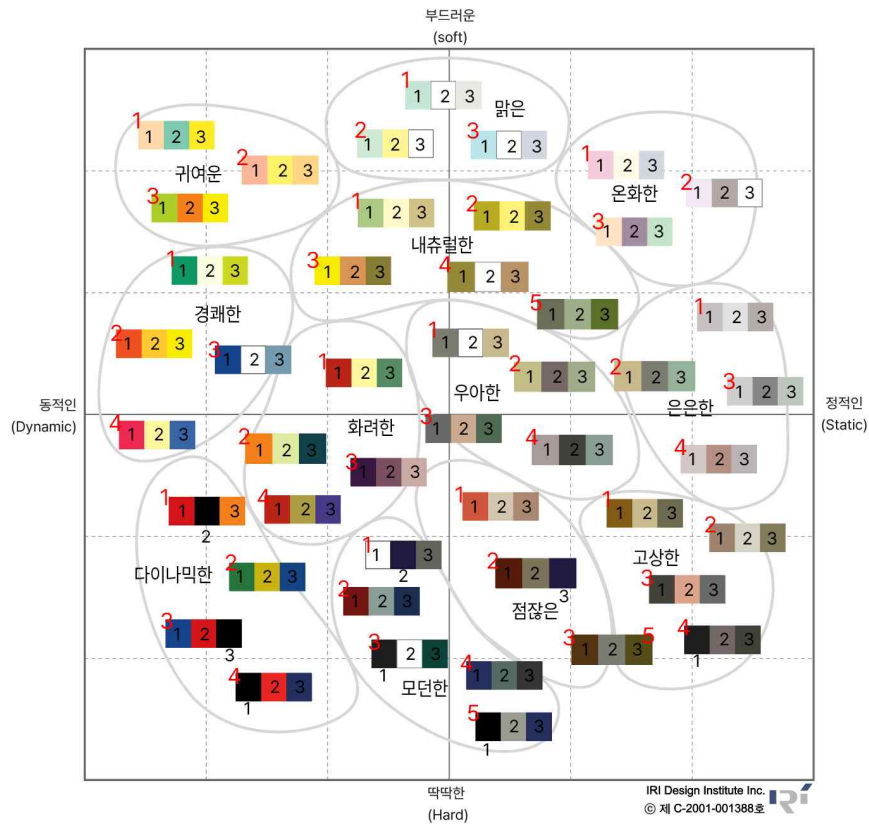
사용자 입력 텍스트	감정 분류 결과		인공지능 맥락 해석
	주 감정	보조 감정	이미지 형용사
네 눈망울에서는 초록빛 오월 하이얀 찔레꽃 내음새가 난다. 네 눈망울에는 초롱초롱한 별들의 이야기를 머금었다.	흐뭇함 (귀여움/예쁨)	기쁨	순수한

이러한 결과는 LLM이 시적 맥락이 갖는 정서적 이미지 특성을 형용사 수준에서 재구성할 수 있음을 보여준다. 특히 ‘순수한’이라는 형용사는 흐뭇함(귀여움/예쁨)과 기쁨의 감정을 감각적으로 구체화하여, 시의 정서적 분위기를 색채화 과정으로 연결하는 핵심 매개로 작용한다. 이를 통해 시의 감정은 ‘흐뭇함’, ‘기쁨’과 같은 추상적 범주를 넘어, ‘순수한’, ‘맑은’, ‘포근한’과 같은 이미지 형용사를 통해 더욱 구체적인 시적 이미지와 감각 경험으로 확장될 수 있다.

이렇게 선택된 이미지 형용사 '순수한'과 주 감정 '흐뭇함(귀여움/예쁨)', 보조 감정 '기쁨'은 이후 최종 2색 배색 팔레트의 명칭으로 활용되며, 시의 감정과 이미지를 함께 반영하는 하나의 공동 해석 결과물로 제시된다.

4) 2색 배색 팔레트 선택

I.R.I 이미지 스케일은 앞서 설명한 형용사 이미지 스케일 뿐만 아니라 배색 이미지 스케일도 제공한다(그림 III-4 참고). 특히 배색 이미지 스케일은 12개의 이미지 형용사군에 대응하는 배색 구조를 포함하며, 각 형용사군의 특성에 부합하는 색채 조합 원리를 바탕으로 구성되어 있다. 또한, 배포된 AI 파일을 Adobe Illustrator에서 분석한 결과, 각 배색 팔레트에 대응하는 RGB 색상값과 배색 구조 정보를 확인할 수 있었다. 본 연구는 이를 추출·정규화하여 감정-색채 변환 알고리즘의 색조 조정 단계에 활용했다. 즉, 인공지능이 선택한 이미지 형용사가 속한 형용사군에 대응하는 배색 팔레트를 검색하고, 해당 팔레트의 명도와 채도 정보를 참조하여 주조색과 보조색의 색조를 조정하도록 설계했다.



[그림 III-4] I.R.I 배색 이미지 스케일의 배색 그룹 및 색상 ID 체계

앞선 단계에서 LLM은 신석정의 「네 눈망울에서는」 시 구절에 대해 이미지 형용사 ‘순수한’을 선택했다. ‘순수한’은 ‘온화한’ 형용사군에 속하며(그림 III-3 참고), 그림 III-4를 살펴보면 해당 형용사군에는 총 3개의 배색 팔레트가 포함되어 있다. 감정-색채 변환 알고리즘은 이 가운데 하나의 배색 팔레트를 무작위로 선택한 후, 팔레트 내 인접한 색상 쌍을 추출하여 주조색과 보조색의 색조 조정 기준으로 활용한다.

본 연구는 이러한 배색 정보를 기계가 처리할 수 있도록 총 141개의 배색 팔레트 데이터를 CSV 형식으로 구조화했다. 각 형용사군은 복수의 배색 팔레트(배색그룹_id)로 구성되며, 각 배색 팔레트는 개별 색상 데이터(단일색상_id)로 이루어진다(표 III-7 참고). 해당 데이터는 함께 읽기(Co-Reading) 시스템 코드 저장소에 포함하여 공개했다.

[표 III-7] ‘온화한’ 형용사군의 배색 팔레트 데이터 예시

형용사군	배색그룹_id	단일색상_id	R	G	B
온화한	1	1	237	199	214
온화한	1	2	252	244	224
온화한	1	3	203	209	218
온화한	2	1	236	225	238
온화한	2	2	175	164	162
온화한	2	3	247	247	246
온화한	3	1	252	222	191
온화한	3	2	159	134	157
온화한	3	3	191	222	195

배색은 기본적으로 서로 인접한 색들의 조합을 통해 구현되기 때문에, 감정-색채 변환 알고리즘은 선택된 배색 팔레트에서 인접한 개별 색상 데이터를 무작위로 선택하고, 이를 주조색과 보조색의 색조 조정을 위한 기준값으로 활용한다. 예를 들어, ‘순수한’ 형용사에 대응하는 ‘온화한’ 배색 팔레트 가운데 하나가 선택되어 배색그룹_id가 ‘1’로 결정된 경우, 시스템은 해당 팔레트의 단일색상_id를 기준으로 인접한 색상 쌍을 구성한다. 이때 가능한 조합은 (1,2), (2,3), (2,1), (3,2)이며, 이 가운데 하나의 조합이 무작위로 선택된다. 이후 선택된 두 색상은 순서에 따라 주조색과 보조색에 대응되어 명도와 채도를 조정한다. 이러한 무작위 선택 방식은 동일한 이미지 형용사에 대해서도 다양한 색채 조합과 해석 가능성을 유지하기 위해서다.

단, I.R.I 배색 이미지 스케일에 포함된 검정(rgb(0,0,0))과 하양(rgb(255,255,255))은 각각 명도 0%와 100%에 해당하여, 색조 조정 단계에서 주조색과 보조색을 모두 극단적인 명도값을 갖게 되는 문제가 발생한다. 따라서 해당 색상 데이터들은 배색에 적합하지 않은 경우로 판단하여, 색상 선택 단계에서 제외했다. 예를 들어, 단일 색상 id ‘1’이 검정 또는 하양인 경우에는 (2,3)과 (3,2) 조합 중에서만 선택되며, 단일 색상 id ‘3’이 검정 또는 하양이면 (1,2)와 (2,1), 단일 색상 id ‘2’가 검정 또는 하양이면 (1,3)과 (3,1) 조합 가운데서만 선택되도록 설계했다. 반면, 명도 0%와 100%에 해당하지 않는 회색 계열 색상은 배색에 그대로 활용했다.

5) 색조 조정 및 시각화

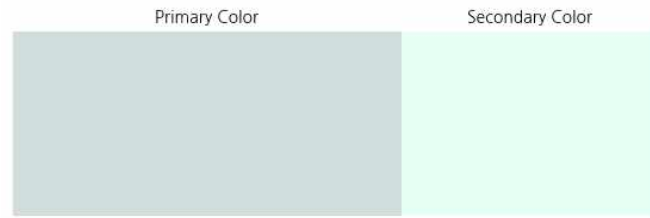
앞선 단계에서 신석정의 시 구절은 주조색 청록(BG)과 보조색 초록(G)으로 결정됐다. 또한, 컨텍스트 엔지니어링을 통해 이미지 형용사 ‘순수한’을 도출됐으며, 이에 대응하는 ‘온화한’ 형용사군의 배색 팔레트 가운데 배색그룹_id ‘1’이 선택됐다.

함께 읽기(Co-Reading) 시스템은 감정-색채 알고리즘의 색조 조정 규칙에 따라 주조색과 보조색의 RGB 값을 HSB 색체계로 변환한 뒤, 선택된 배색 팔레트의 명도와 채도 정보를 반영하여 색조를 조정한다. 표 III-8은 신석정의 「네 눈망울에서는」에 적용된 RGB-HSB 변환 및 색조 조정 결과를 나타낸 것이다.

[표 III-8] 신석정 사례의 주조색과 보조색 색조 조정 결과

조정 단계	주조색		보조색	
	RGB	HSB	RGB	HSB
색조 조정 전	R: 0 G: 151 B: 133	H: 173° S: 100% B: 59%	R: 0 G: 116 B: 56	H: 149° S: 100% B: 45%
색조 조정 후	R: 203 G: 216 B: 215	H: 173° S: 6% B: 85%	R: 224 G: 249 B: 236	H: 149° S: 10% B: 98%

조정된 HSB 값은 다시 RGB 색상값으로 변환되며, 주조색과 보조색이 6:4 비율로 구성된 2색 배색 팔레트가 출력된다. 이와 같은 과정을 통해 함께 읽기(Co-Reading) 시스템은 감정 분류 결과로부터 도출된 색상 정보를 유지하면서도, LLM이 해석한 이미지 형용사를 반영하여 시의 정서와 이미지를 함께 담아내는 최종 색채 팔레트를 생성한다(그림 III-5 참고).



순수한, 흐릿함(귀여움/예쁨)과(와) 기쁨



감각적인, 흐릿함(귀여움/예쁨)과(와) 기쁨

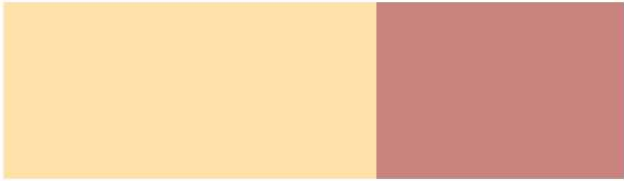
[그림 III-5] 신석정, 「네 눈망울에서는」
최종 시각화 2색 배색 팔레트

그림 III-5에서 보이는 것 처럼, LLM 이미지 형용사 선택 과정에서는 앞서 제시한 ‘순수한’ 외에도 ‘감각적인’이 선택된 사례가 확인됐다. 이는 동일한 청록과 초록의 조합이라 하더라도, 인공지능이 해석한 이미지 형용사에 따라 명도와 채도가 다르게 조정되어 색다른 감각적 분위기의 색채 팔레트가 생성될 수 있음을 시사한다. 이는 함께 읽기(Co-Reading) 시스템이 시 텍스트에 대한 인공지능의 해석 결과를 핵심 매개로 활용하여 색채를 동적으로 구성하고 있음을 보여준다.

이처럼 함께 읽기(Co-Reading) 시스템은 동일한 시 구절에 대해서도 다양한 색채 표현이 가능하다. 이는 감정-색채 변환 알고리즘을 통한 정량적 일관성을 유지하면서도, 인공지능의 해석을 반영한 시각적 다양성과 심미성을 동시에 확보하기 위한 전략이다. 또한, 생성된 색채 팔레트를 ‘순수한, 흐릿함(귀여움/예쁨)과 기쁨’와 같이 이미지 형용사와 감정 이름이 결합된 명칭으로 제시함으로써, 사용자가 시의 정서와 이미지를 공감각적으로 해석할 수 있도록 유도한다.

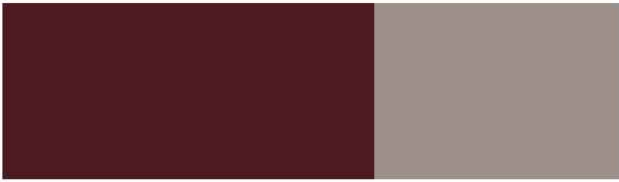
다음 표 III-9a와 표 III-9b는 시인 백석과 김수영의 시 텍스트에 대해 함께 읽기를 실행한 결과를 나타낸 것이다

[표 III-9a] 백석(2023), 「나와 나타샤와 흰 당나귀」⁶³⁾ 함께 읽기 결과

입력한 시 텍스트	감정분류 결과 (상위 5개)	인공지능의 이미지 형용사 출력
눈은 푹푹 나리고 아름다운 나타샤는 나를 사랑하고 어데서 흰 당나귀도 오늘밤이 좋아서 응앙응앙 울을 것이다	흐뭇함(귀여움/예쁨): 0.95 아껴주는: 0.93 기쁨: 0.90 행복: 0.80 기대감: 0.80	사랑스러운
	주 감정	주조색
	흐뭇함(귀여움/예쁨)	주황
	보조 감정	보조색
	아껴주는	빨강
2색 배색 팔레트 변환		
Primary Color Secondary Color  사랑스러운, 흐뭇함(귀여움/예쁨)과(와) 아껴주는		

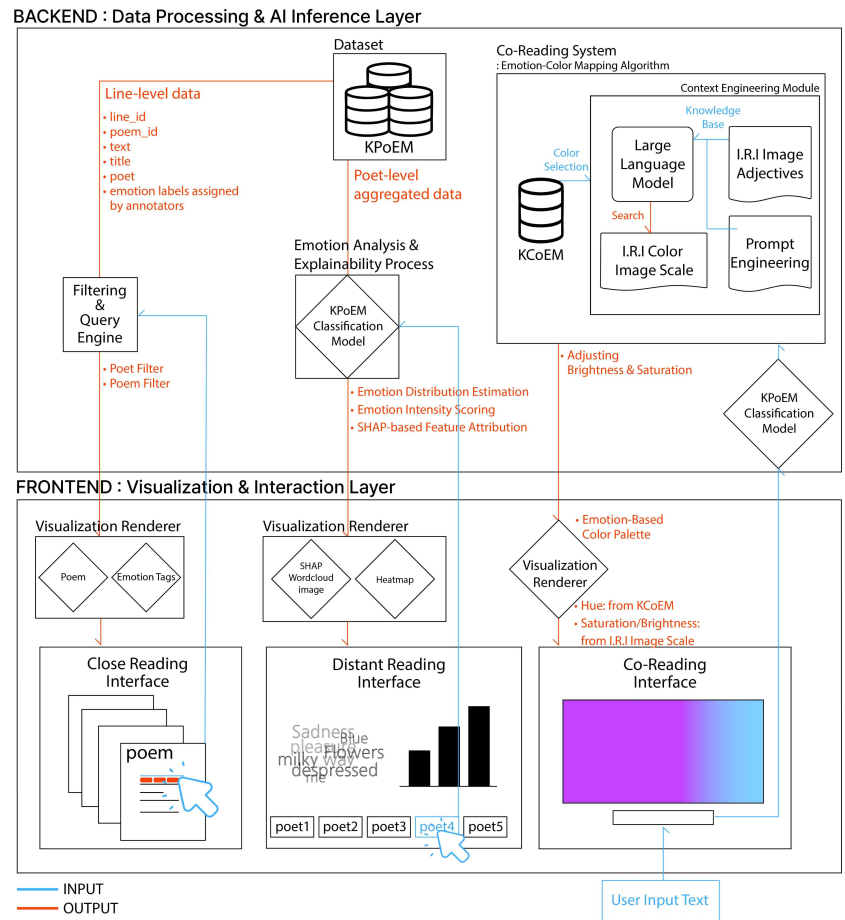
63) 백석. (2023). *백석 전 시집: 나와 나타샤와 흰 당나귀*. 스타북스.

[표 III-9b] 김수영(1974), 「푸른 하늘을」⁶⁴⁾ 함께 읽기 결과

입력한 시 텍스트	감정분류 결과 (상위 5개)	인공지능의 이미지 형용사 출력
자유를 위해서 비상하여 본 일이 있는 사람이면 알지. 노고지리가 무엇을 보고 노래하는가를 어째서 자유에는 피의 냄새가 섞여있는가를 혁명은 왜 고독한 것인가를	안타까움/실망: 0.94 슬픔: 0.88 깨달음: 0.81 서러움: 0.81 불안/걱정: 0.81	깊은
	주 감정	주조색
	안타까움/실망	빨강
	보조 감정	보조색
	슬픔	주황
2색 배색 팔레트 변환		
Primary Color Secondary Color  깊은, 안타까움/실망과(와) 슬픔		

64) 김수영. (1974). *거대한 뿌리*. 민음사.

3. 웹 인터페이스 구현 및 결과⁶⁵⁾



[그림 III-6] KPoEM 인터페이스 프로토타입 구조 설계 개념도

본 연구가 제안하는 함께 읽기(Co-Reading) 시스템은 인간과 인공지능이 데이터를 매개로 문학 텍스트를 함께 해석하고, 그 결과를 색채 이미지의 형태로 공유하는 커

65) 본 연구에서 구현한 웹 인터페이스의 소스코드는 GitHub 저장소에서 확인할 수 있다. 해당 저장소에는 웹 인터페이스 구현 코드뿐만 아니라, 멀리서 읽기 인터페이스에서 활용한 분석 코드와 분석에 사용된 데이터 파일이 함께 포함되어 있다.

임이로. (2026). *KPoEMInterface* [Software]. GitHub.
<https://github.com/poet-developer/KPoEMInterface>

또한 실제 웹 인터페이스 프로토타입은 GitHub Pages를 통해 공개했다.
 임이로. (2026). *KPoEMInterface* [Web app]. GitHub Pages.
<https://poet-developer.github.io/KPoEMInterface>

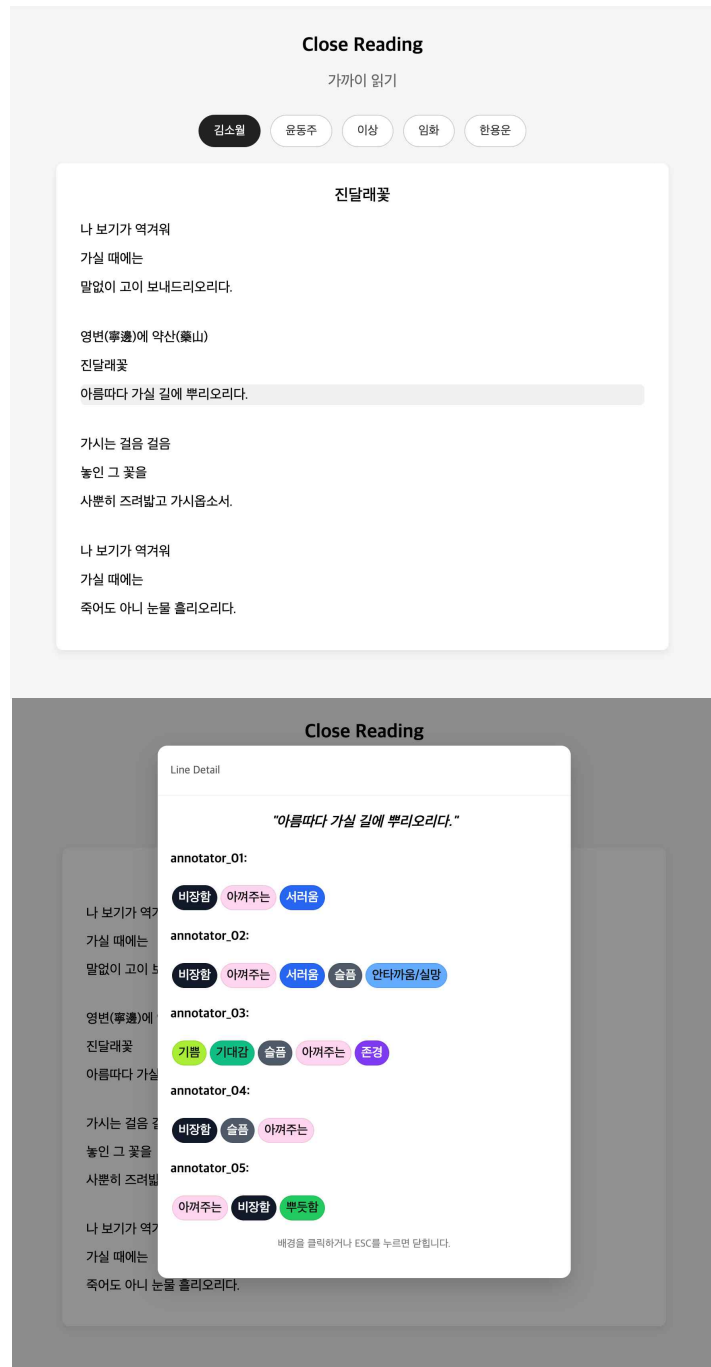
뮤니케이션 구조를 지향한다. 이러한 구조는 함께 읽기의 결과를 사용자에게 다시 제시하여 탐색과 재해석을 가능하게 하는 상호작용(interaction)을 전제로 하며, 이는 웹 인터페이스를 통해 구현된다. 인터페이스는 사용자의 사고와 해석을 매개하는 공간이며(Drucker, 2014), 데이터는 인터페이스를 통해 동적으로 조직되고 경험될 수 있기 때문이다(Manovich, 2001). 따라서 본 연구가 구현한 웹 인터페이스는 다양한 데이터 시각화 방식을 통해 감정 데이터를 독자의 읽기 과정에서 경험하는 상호작용적 읽기 공간이다.

그림 III-6는 본 연구에서 제안하는 KPoEM 데이터셋 기반 감정 시각화 인터페이스의 개념적 구조를 도식화한 것이다. 본 연구에서는 가까이 읽기(close reading), 멀리서 읽기(distant reading), 함께 읽기(Co-Reading) 모듈 간의 관계와 사용자 경험 흐름을 중심으로 정적(static) 프로토타입을 구현했다. 이를 통해 각 읽기 방식의 구성 요소와 상호작용 구조를 시각적으로 제시하고자 한다.

KPoEM 인터페이스는 기존의 정량적 통계 중심 데이터 시각화를 넘어, 함께 읽기를 통해 시-감정-색채가 연결되는 해석 과정을 탐색할 수 있도록 설계했다. 사용자는 인터페이스를 통해 감정 데이터를 다양한 관점에서 살펴보고, 시의 의미를 능동적으로 구성하는 읽기 경험에 참여할 수 있다.

본 장에서는 인터페이스의 구체적인 작동 방식과 사용자 경험 측면을 설명하기 위해, KPoEM 데이터셋 내에서 가장 많은 작품 수를 포함하고 있는 시인 김소월을 중심 사례로 제시한다. 한편, 윤동주, 이상, 임화, 한용운 등 나머지 네 시인에 대한 가까이 읽기, 멀리서 읽기, 그리고 함께 읽기 인터페이스 구현 결과 역시 배포된 웹 페이지를 통해 확인할 수 있다.

1) 가까이 읽기 인터페이스 : 김소월, 「진달래꽃」



[그림 III-7] 김소월, 「진달래꽃」 가까이 읽기 인터페이스 예시

가까이 읽기 인터페이스는 KPoEM 데이터셋의 행 단위 감정 주석 데이터를 정밀하게 탐색하기 위해 설계한 인터페이스로, 시 텍스트에 나타나는 미시적 정서 표현을 시각적으로 드러내는 데 초점을 둔다(그림 III-7 참고). 사용자는 시인과 작품을 선택한 뒤, 전체 시 텍스트를 행 단위로 탐색할 수 있으며, 특정 행을 선택하면 해당 행에 대한 주석자들의 감정 태그를 모달(modal) 형태로 확인할 수 있다. 이러한 구조는 전통적인 가까이 읽기에서 강조되어 온 세밀한 텍스트 해석 방식을 인터페이스로 구현한 것이다. 사용자는 가까이 읽기에 참여함으로써, 시적 의미가 형성되는 최소 단위에서 주석자 간 해석의 차이 및 중첩 양상을 탐색할 수 있도록 한다. 이를 통해 가까이 읽기 인터페이스는 정적인 시 텍스트를 인터랙티브 객체로 전환한다. 이때, KPoEM 주석 데이터는 개별 감정의 단일한 결과값이 아니라, 시구를 읽는 과정에서 축적된 다중 해석의 기록으로 이해될 수 있다.

가까이 읽기 인터페이스는 크게 1) 전체 텍스트 제시 영역, 2) 행 단위 클릭 인터랙션, 3) 감정 주석 모달로 구분된다. 각 요소는 사용자가 시 텍스트를 단계적으로 읽고, 감정 해석의 차이를 탐색할 수 있도록 유기적으로 결합되어 있다.

① 전체 텍스트 제시 영역

전체 텍스트 제시 영역은 시를 하나의 완결된 텍스트로 제시하여, 사용자가 작품 전체를 정독할 수 있도록 구성했다. UI(User Interface) 측면에서 시 텍스트는 화면 중앙에 배치되며, 충분한 여백과 간격을 통해 행 단위의 구분이 명확하게 드러나도록 설계했다. 특히 행 간격과 타이포그래피의 조정은 시의 리듬과 호흡을 시각적으로 보존하는 동시에, 이후 인터랙션의 기본 단위가 되는 개별 행을 자연스럽게 강조하는 역할을 한다.

② 행 단위 클릭 인터랙션

행 단위 클릭 인터랙션은 사용자의 해석 행위를 촉발하는 핵심 경험 요소다. 사용자가 특정 행 위에 마우스를 올리거나 클릭할 경우, 해당 행은 hover 이벤트를 통해 시각적으로 강조되어 선택 가능한 상태임을 드러낸다. 이는 텍스트를 단지 읽는 대상에서 ‘탐색 가능한 객체’로 전환시키는 장치다. 사용자는 특정 행을 선택함으로써 그 부분에 대한 감정 데이터를 호출한다. 이러한 구조는 전통적인 가까이 읽기의 방식, 즉 특정 구절을 반복적으로 읽고 집중하는 방식을 디지털 환경에서 재현한 것이

다. 다시 말해, 클릭이라는 간단한 행위는 ‘이 행을 자세히 읽겠다’는 해석적 의도를 표현하는 행위로 기능한다.

③ 감정 주석 모달

감정 주석 모달은 선택된 행에 대한 KPoEM 데이터셋 주석자들의 감정 해석을 집약적으로 제시하는 공간이다. 사용자가 특정 행을 선택하면 화면 중앙에 모달 창이 나타나며, 해당 시구가 다시 한 번 강조된 형태로 모달 상단에 제시된다. 이때 모달 외부 배경 영역을 어둡게 처리함으로써, 사용자의 시선이 선택된 행과 주석 정보에 집중되도록 유도한다. 이러한 UI 설계는 정보의 위계 구조를 명확하게 하여, 현재 사용자가 탐색하고 있는 해석 대상을 직관적으로 인식할 수 있도록 한다.

모달 내부에서는 KPoEM 데이터셋의 다섯 명 주석자가 부여한 감정 라벨이 각각 독립된 영역에 배치된다. 즉, 사용자는 하나의 정답을 확인하는 것이 아니라 서로 다른 해석들이 병렬적으로 놓인 상태를 탐색하게 된다. 이러한 다중 해석 비교 구조는 감정 라벨을 통합하거나 평균화하지 않고, 각 주석자의 해석 결과를 그대로 보여줌으로써 해석의 차이를 제거하지 않는다. 이때 사용자는 클릭한 행에서 어떠한 감정이 반복되는지, 그리고 어떠한 감정이 다르게 해석되는지를 직접 비교할 수 있다. 예를 들어, 동일한 행에 대해 여러 주석자가 ‘아껴주는’이라는 감정을 공유하는 경우, 동일 색상의 태그가 반복적으로 나타나면서, 해석의 공통점이 시각적으로 강조된다. 반면 서로 다른 감정이 제시될 경우, 상이한 색상의 태그가 병렬적으로 배열되어 해석의 분산과 다양성이 직관적으로 드러난다. 이는 인터페이스가 하나의 해석을 제시하는 것이 아니라, 사용자가 다양한 해석의 관계를 탐색하고 스스로 의미를 구성하도록 유도하는 구조라고 할 수 있다.

또한, 모달 창은 배경 영역을 클릭하거나 ESC 키를 입력하면 닫히도록 설계하여, 사용자가 특정 행에 대한 독해를 마친 뒤 자연스럽게 전체 텍스트 제시 영역으로 복귀할 수 있도록 했다. 이러한 선택과 복귀의 순환 구조는 사용자가 여러 행을 반복적으로 비교하고 탐색하는 가까이 읽기의 경험을 디지털 환경에서 구현한 것이다.

2) 멀리서 읽기 인터페이스 : 김소월 전체 작품 통계



[그림 III-8] 김소월 전체 작품의 멀리서 읽기 인터페이스 예시

멀리서 읽기 인터페이스는 가까이 읽기의 세밀한 문학 분석을 보완하여, 각 시인의 작품 전체에서 나타나는 감정 분포와 어휘 경향을 거시적으로 조망하기 위한 시각화 인터페이스다(그림 III-8 참고). 본 인터페이스는 KPoEM 데이터셋을 구성하는 KOTE 44개 감정 범주를 기반으로, 특정 시인에게 반복적으로 나타나는 감정의 양상과, 해당 감정이 어떠한 언어적 요소를 통해 형성되는지를 통계적·시각적으로 탐색할 수 있도록 구성했다.

멀리서 읽기 인터페이스는 크게 1) SHAP 기반 워드클라우드 시각화, 2) 시인 단위 감정 분포 그래프, 3) 시인의 작품 단위 감정 히트맵으로 구성된다. 각 시각화 요소는 단어 수준의 감정 기여도 분석에서 출발하여, 시인 및 작품 단위의 감정 분포와 감정 강도를 단계적으로 탐색할 수 있도록 설계했다. 또한, 본 연구의 분석 코드 및 산출물은 연구 재현성을 위해 KPoEMInterface 코드 저장소에 포함하여 공개했다.

① SHAP(Shapley Additive exPlanations) 기반 워드클라우드 시각화

사용자는 감정 범주 선택 UI를 통해 분석하고자 하는 감정을 선택할 수 있으며, 이에 따라 해당 감정의 예측에 긍정적으로 기여한 핵심 어휘들이 SHAP 기반 워드클라우드로 시각화된다. SHAP(v0.51.0)⁶⁶⁾는 머신러닝 모델의 예측 결과에 대해 각 입력 변수가 특정 예측에 얼마나 기여했는지를 정량적으로 산출하는 설명 가능 인공지능(Explainable AI, XAI) 기법이다. 본 연구에서는 이 점을 착안하여 KPoEM 감정 분류 모델이 특정 감정을 예측할 때 중요하게 학습한 어휘를 역으로 추적했다.

멀리서 읽기 인터페이스에서는 선택된 감정의 예측 확률을 증가시키는 방향으로 작용한 어휘, 즉 양(+)의 SHAP 값⁶⁷⁾만을 추출하여 워드클라우드(v1.9.6)⁶⁸⁾로 시각화했다. 이때 단어의 크기는 해당 감정 예측에 대한 기여도를 의미하며, 클수록 감정 예측에 더 큰 영향을 준 어휘를 나타낸다. 또한 조사와 같은 기능어로 인한 시각적 노이즈를 줄이고 의미 있는 어휘 패턴을 강조하기 위해, 워드클라우드 생성 시 2음절 이상의 토큰만을 시각화 대상으로 제한했다.

66) SHAP 라이브러리 관련 정보는 다음 저장소에서 확인할 수 있다.

SHAP. (2026). *SHAP* (v.0.51.0) [Software]. GitHub. <https://github.com/shap/shap>

67) 본 연구에서는 양의 SHAP 값 가운데 0.05 이상의 값을 갖는 어휘만 시각화 대상으로 사용했다. 해당 임계값(threshold)은 시각화 과정에서 지나치게 작은 기여도를 제거하고, 감정과의 관련성이 높은 어휘를 보다 명확하게 제시하기 위해 연구자가 시각화 결과를 반복적으로 검토하는 과정에서 설정했다.

68) 워드클라우드 라이브러리 관련 정보는 다음 웹 페이지에서 확인할 수 있다.

Mueller, A. (2026). *wordcloud* (v1.9.6) [Software]. GitHub. https://github.com/amueller/word_cloud

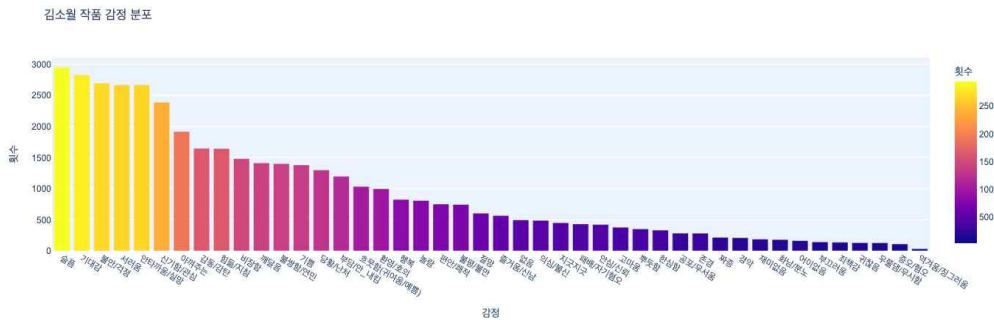


[그림 III-9] 김소월 작품의 SHAP 기반 워드클라우드
(상: 고마움, 하: 죄책감)

이러한 시각화 방식은 감정 형성에 관여하는 시인의 핵심 어휘를 시각적으로 부각시켜, 특정 감정이 어떠한 언어적 맥락 속에서 구성되는지를 직관적으로 드러낸다.

그림 III-9는 김소월 시 작품 전체를 대상으로 SHAP 분석 결과를 시각화한 예시다. ‘고마움’ 감정 범주에서는 ‘사랑’, ‘그대’, ‘고향’, ‘우리’와 같은 어휘가 해당 감정 예측 확률을 높이는 주요 어휘임 확인할 수 있다. 반면, ‘죄책감’ 감정 범주에서는 ‘나는’, ‘물렸’, ‘잊고’와 같은 어휘가 해당 감정 예측에 중요한 영향을 미치는 핵심 어휘로 나타난다. 이는 특정 감정이 시인이 반복적으로 선택하는 어휘와 표현 양식 속에서 구축된다는 점을 보여준다. 따라서 SHAP 기반 워드클라우드는 시인의 언어적 습관과 정서적 표현간의 관계를 시각적으로 해석할 수 있도록 지원한다.

② 시인 단위 감정 분포 그래프



[그림 III-10] 김소월 작품 감정 분포 UI

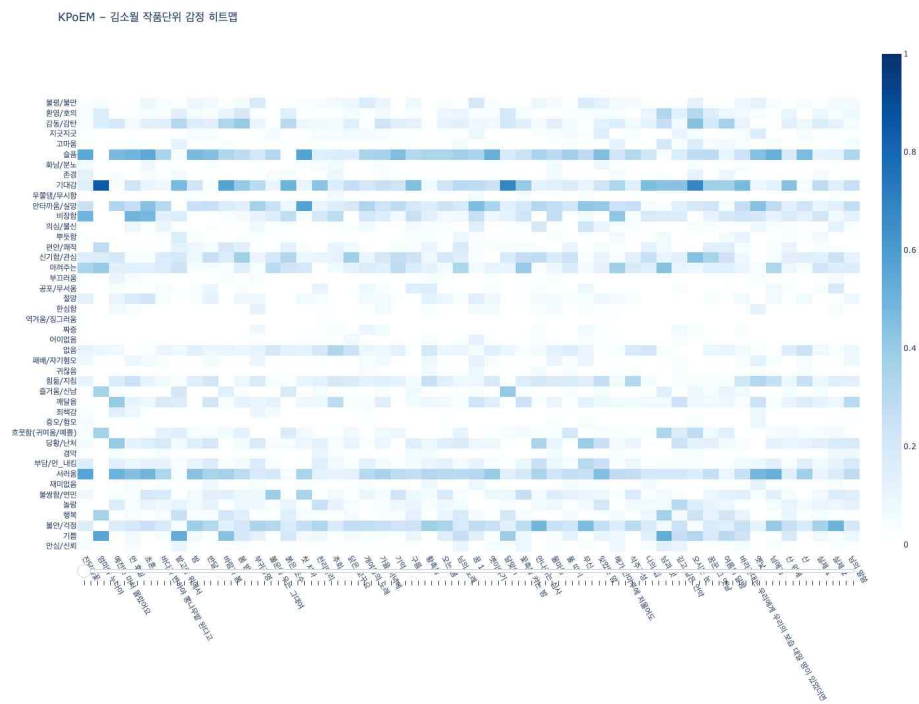
시인 단위 감정 분포 그래프는 특정 시인 작품 전체에 나타나는 감정 분포를 집계하여 시각화함으로써, KPoEM 데이터셋에 기반한 시인의 정서적 경향과 감정 구조를 거시적으로 파악할 수 있도록 한다(그림 III-10 참고). 본 시각화는 Python 기반 시각화 라이브러리인 Plotly(v6.8.0)를 활용하여 이 중 Plotly Express⁶⁹⁾ 모듈을 활용하여 구현했으며, 각 막대는 마우스 이벤트를 통해 해당 감정의 정확한 빈도값을 확인할 수 있도록 설계했다. 또한 감정 빈도에 따라 색상 강도가 변화하도록 구성하여, 감정의 상대적 비중을 직관적으로 해석할 수 있도록 했다. 이때 인터페이스는 Plotly Express가 제공하는 HTML 기반 인터랙티브 시각화 기능을 별도의 수정 없이 그대로 활용했다.

이러한 시인 단위 감정 분포 그래프는 특정 시인의 정서적 중심축을 드러낸다. 사용자는 이를 통해 어떤 감정이 상대적으로 많이 또는 적게 나타나는지를 직관적으로 비교할 수 있으며, 나아가 시인의 작품 전반에 걸쳐 나타나는 감정 구조와 정서적 특성을 한눈에 파악할 수 있다. 김소월의 시 작품은 전체적으로 슬픔과 기대감, 불안/걱정, 서러움, 안타까움/실망 등의 감정이 높은 빈도로 나타난다. 이는 김소월의 시 세계가 상실과 그리움의 정서를 중심으로 하면서도, 그 안에 희망과 기다림의 감정이 복합적으로 공존하는 정서 구조를 형성하고 있음을 보여준다.

69) Plotly Express는 Python 환경에서 인터랙티브 웹 기반 시각화를 생성하기 위한 Plotly의 데이터 시각화 모듈이다. Plotly 라이브러리 관련 정보는 다음 웹 페이지에서 확인할 수 있다.

Plotly. (2026). *Plotly.py* (v6.8.0) [Software]. GitHub.
<https://github.com/plotly/plotly.py>

③ 시인의 작품 단위 감정 히트맵



[그림 III-11] 김소월 작품-감정 히트맵 UI 예시

시인의 작품 단위 감정 히트맵(heatmap)은 시인의 작품별 44개 감정 범주의 분포를 행렬 형태로 시각화한 것이다(그림 III-11 참고). 히트맵의 각 셀은 특정 작품에서 나타나는 감정의 상대적 강도를 색상 농도로 표현하며, 이를 통해 작품 간 감정 밀도의 차이, 특정 감정의 집중 양상, 그리고 시인별 정서 구조의 패턴을 한눈에 비교할 수 있다.

각 작품의 감정값은 KPOEM 데이터셋에 부여된 멀티라벨 감정 주석을 기반으로 산출했다. 먼저, 각 행에 대해 주석자가 부여한 복수의 감정 라벨을 44차원의 원-핫(one-hot) 인코딩을 통해 벡터로 변환한 뒤, 동일 작품에 속한 모든 데이터에 대해 평균값을 계산했다. 이후, 감정 주석자간 평균을 추가로 적용함으로써 최종적으로 작품 단위의 감정 분포를 산출했다. 이때 각 차원의 값은 특정 감정이 작품 전반에 걸쳐 나타나는 상대적 빈도 및 강도를 나타낸다. 히트맵에서는 이러한 감정 분포를 색상 농도로 대응시켜 시각화했으며, Plotly(v6.8.0) 기반 인터랙티브 시각화를 통해 각 셀에 마우스를 올리면 작품명, 감정 이름, 감정값이 즉시 표시되도록 구성했다.

또한, 슬라이더를 기반 화면 전환 UI을 추가하여 많은 작품을 효율적으로 탐색할 수 있도록 구성했으며, 이를 통해 사용자의 인지 부담을 줄이고 탐색 흐름을 유지하도록 했다. 사용자는 이를 통해 특정 감정이 어떤 작품에서 집중되는지, 그리고 어떤 감정이 반복적으로 등장하는지를 비교함으로써, 시인의 정서 구조를 거시적으로 탐색할 수 있다.

3) 함께 읽기 인터페이스 : 김소월, 「진달래꽃」



[그림 III-12] 김소월, 「진달래꽃」 함께 읽기 인터페이스 예시

함께 읽기(Co-Reading) 인터페이스는 감정을 고정된 라벨이나 수치로 환원하지 않고, 인간과 인공지능의 해석 결과를 색채 이미지로 제시함으로써, 시 읽기를 언어와 색채가 교차하는 공동 해석의 과정으로 재구성한다(그림 III-12 참고). 이때 사용자에게 반환되는 것은 정답이나 분석 결과가 아니라, 인간과 인공지능이 함께 시를 읽고 생성한 해석의 흔적으로서의 감정-색채 팔레트다.

현재 프로토타입에서 함께 읽기 인터페이스는 크게 1) 작품 선택 영역, 2) 텍스트 확인 영역, 3) 실행 및 상태 인식 영역, 4) 색채 결과 시각화 영역, 5) 색상 정보 확인 영역으로 구성된다. 이러한 구성 요소는 사용자가 함께 읽기 시스템의 처리 구조를 직관적으로 이해할 수 있도록, 상단에서 하단으로 이어지는 독해 흐름에 따라 수직으로 배열했다. 이는 사용자가 시를 먼저 읽고, 해석을 요청하며, 결과가 드러나는 시간을 경험한 뒤, 최종적으로 감정과 색채를 함께 해석하도록 유도하기 위한 설계다.

① 작품 선택 영역

작품 선택 영역은 함께 읽기 인터페이스의 출발점으로, 사용자가 함께 읽기 대상 텍스트를 선택하는 단계이다. 현재 구현된 프로토타입에서는 자유 입력 방식 대신 드롭다운(select) 기반의 작품 선택 방식을 적용했다. 이는 정적 프로토타입 단계에서, 시스템의 구조를 명확하게 드러내고 사용 흐름을 안정적으로 제어하기 위한 설계다. 사용자는 ‘시 선택’ 항목에서 작품을 고르며, 선택된 작품은 즉시 아래의 텍스트 영역에 반영된다.

② 텍스트 확인 영역

텍스트 확인 영역은 선택된 시 본문이 나타나는 부분으로, 사용자가 현재 해석의 대상이 되는 텍스트를 충분히 인지할 수 있도록 돕는다. 이 영역은 읽기 전용 텍스트 형태로 제공되어, 사용자가 작품을 선택하면 시 본문이 충분한 크기의 영역 안에 표시된다. 이 텍스트 영역은 공동 해석을 위한 전제 조건으로서, 인간 독해의 공간으로 기능한다.

③ 실행 및 상태 인식 영역

실행 및 상태 인식 영역은 사용자가 함께 읽기(Co-Reading)를 요청하고, 시스템이

이에 응답하는 과정을 경험하도록 설계한 영역이다. 이는 ‘함께 읽기’ 버튼과 ‘지우기’ 버튼, 그리고 ‘Reading...’ 상태 표시 요소로 구성된다. 먼저 ‘함께 읽기’ 버튼은 사용자가 시를 선택한 뒤 시-감정-색채 변환을 요청하는 시작점이며, ‘지우기’ 버튼은 선택과 결과를 초기 상태로 되돌려 사용자가 결과를 반복적으로 비교하고 탐색할 수 있도록 한다.

또한 ‘Reading...’ 애니메이션은 함께 읽기 시스템이 시를 읽고 있다는 감각을 형성하는 역할을 수행한다. 기술적으로는 단순한 로딩(loading) 표시 기능이지만, 짧은 대기 시간을 통해 사용자가 결과를 즉시 소비하기보다 함께 읽기가 진행되는 과정을 경험하도록 유도한다. 이는 결과의 생성을 인간과 인공지능이 함께 시를 읽고 해석하는 시간적 과정으로 인식하게 하는 장치로 기능한다.

④ 색채 결과 시각화 영역

색채 결과 시각화 영역은 함께 읽기 인터페이스의 핵심 시각화 영역이다. 최종적으로 생성된 2색 배색 팔레트가 하나의 넓은 면으로 제시된다. 본 연구에서는 색채 정보를 데이터만으로 제시하기보다, 색채가 형성하는 정서적 분위기와 감각적 인상을 사용자가 직접 경험할 수 있도록 비교적 큰 면적의 색면(色面)으로 구성했다.

이 영역의 UI는 감정의 복합성을 시각적으로 표현하기 위한 구조를 가진다. 함께 읽기 시스템은 주조색과 보조색으로 구성된 2색 배색 팔레트를 출력하며, 이는 하나의 시가 단일한 감정으로 환원되지 않는다는 점을 반영한다. 예를 들어 김소월의 「진달래꽃」에서 나타난 ‘슬픔’과 ‘서러움’의 조합은 서로 중첩된 정서와 분위기를 형성하며, 본 인터페이스는 이를 색채의 조합과 그라데이션을 통해 시각적으로 표현한다.

또한 인터페이스는 감정 결과와 함께 인공지능이 선택한 이미지 형용사를 함께 제시한다. 이를 통해 사용자는 색채를 지각하는 데 그치지 않고, 색면 상단에 제시된 2색 배색 팔레트의 명칭인 ‘품위있는, 슬픔과 서러움’과 같이, 언어와 색채가 결합된 공동 해석 결과물을 새로운 감각적 표현으로 경험하게 된다. 특히 이미지 형용사와 감정의 조합은 기존의 감정 범주만으로는 포착하기 어려운 분위기와 심상을 구체화하며, 시 텍스트에 대한 새로운 연상과 해석 가능성을 제공한다. 이러한 과정은 색채 이미지를 다시 언어적 표현으로 환원하고 재해석하는 순환 구조를 형성함으로써,

인간과 인공지능이 함께 의미를 구성하는 함께 읽기(Co-Reading)의 경험을 확장한다.

⑤ 색상 정보 확인 영역

색상 정보 확인 영역은 앞서 출력된 색면과 기계 가독형 데이터를 연결하는 영역이다. 출력된 팔레트 하단에는 각 색상에 대응하는 RGB 값과 HEX 값이 함께 제시된다. 이는 시-감정-색채 변환 결과를 색면이라는 미적 경험으로 제시할 뿐만 아니라, 사용자에게 참조 가능한 색채 데이터를 함께 제공하기 위한 UI다. 사용자는 상단에 출력된 색면을 통해 먼저 색채를 감각적으로 경험하고, 하단의 색상값을 통해 그 결과를 구체적이고 기술적인 수준에서 확인할 수 있다.

이와 같이 함께 읽기 인터페이스는 감각적 경험과 계산적 정보를 분리하지 않고, 하나의 연속된 해석 결과 구조 안에서 함께 제시한다. 이를 통해 사용자는 공동 해석의 결과를 다시 데이터의 관점에서 참조하고 활용할 수 있다.

4) 가까이 읽기, 멀리서 읽기, 함께 읽기 인터페이스 비교

표 III-10는 본 연구에서 구현한 가까이 읽기, 멀리서 읽기, 그리고 함께 읽기 인터페이스의 기능과 목적, 그리고 해석 방식의 차이를 비교하여 정리한 것이다.

[표 III-10] 가까이 읽기, 멀리서 읽기, 함께 읽기 인터페이스 비교

구분	가까이 읽기 (Close Reading)	멀리서 읽기 (Distant Reading)	함께 읽기 (Co-Reading)
분석 대상	시의 행 단위	시인 및 작품 단위	사용자가 선택한 시 텍스트
주요목적	미시적 감정 표현과 다중 해석 구조 탐색	시인의 감정 분포와 언어적 경향의 거시적 구조 탐색	인간-AI 공동 해석 기반 감정-색채 시각화
핵심 데이터	KPoEM 행 단위 감정 주석 데이터 (5인 주석)	KPoEM 통계 데이터 및 SHAP 기여 어휘	감정 분류 결과, KCoEM, 이미지 형용사 및 색채 변환 데이터
시각화 방식	행 단위 인터랙션 및 감정 태그	SHAP 워드클라우드, 통계 그래프 및 히트맵	2색 배색 팔레트

사용자 경험	특정 시구 감정 해석 탐색	시인 감정 패턴 거시적 탐색	시-감정-색채 기반 공동 해석 경험
해석의 성격	해석의 다양성과 중층성 강조	통계적 패턴과 구조적 경향 분석	인간-AI 상호작용을 통한 감정의 감각적 재구성
인공지능 의 역할	없음	KPoEM 감정 분류 모델의 예측 근거 설명	공동 독자(co-reader)로서 시 해석과 의미 구성에 참여

먼저 가까이 읽기 인터페이스는 시의 행 단위를 중심으로, 주석자 간 감정 해석의 차이와 중층성을 탐색하는 데 초점을 둔다. 이는 개별 텍스트에 대한 미시적 접근을 기반으로 하며, KPoEM 감정 주석 데이터를 통해 해석의 다양성을 드러내는 전통적인 가까이 읽기 방식을 웹 인터페이스로 구현한 것이다.

반면, 멀리서 읽기 인터페이스는 시인 및 작품 단위를 대상으로 감정 분포와 언어적 경향을 통계적으로 시각화함으로써, 시인의 감정 패턴과 구조적 경향을 파악하는 데 목적이 있다. 이는 개별 텍스트의 의미를 넘어 데이터 집합 전체를 하나의 분석 단위로 간주하는 거시적 독해 방식으로, 감정과 어휘의 분포를 통해 시인의 언어적·정서적 경향을 탐색하는 정량·통계적 접근 방식을 보여준다.

마지막으로 함께 읽기(Co-Reading) 인터페이스는 사용자가 선택한 시 텍스트를 기반으로 인간과 인공지능이 공동으로 시를 해석하고, 그 결과를 색채로 변환하는 상호작용적 독해 환경을 구성한다. 이는 감정을 고정된 분석 결과로 제시하는 대신, 감정-색채-언어가 결합된 2색 배색 팔레트를 통해 해석 과정 자체를 경험하도록 한다는 점에서 기존의 독해 방식과 구별된다.

결과적으로 본 연구의 인터페이스 구조는 미시-거시-상호작용이라는 세 가지 해석 층위를 통합하는 문학 독해 구조를 제안한다. 특히 함께 읽기 인터페이스는 감정 데이터를 감각적 경험으로 전환함으로써, 문학 텍스트를 인간과 인공지능이 함께 해석하고 의미를 구성할 수 있는 새로운 상호작용적 읽기 환경의 가능성을 제시한다.

IV. 결론

본 연구는 한국 근현대 시의 감정을 색채로 변환하는 과정을 통해 인간과 인공지능이 함께 문학 텍스트를 해석하는 환경(system), 즉 ‘함께 읽기(Co-Reading)’의 가능성을 제시했다. 특히 가까이 읽기와 멀리서 읽기를 비교 기준으로 설정하고, 이를 각각 인터페이스로 구현하여 비교 및 분석함으로써, 함께 읽기가 문학 데이터의 해석적 경험을 어떻게 확장하는지를 실험적으로 탐색했다.

이를 위해 한국 근현대 시-감정 데이터셋 KPoEM과 색채-감정 데이터셋 KCoEM을 구축하고, KPoEM 감정 분류 모델과 감정-색채 변환 알고리즘을 개발하여, 대규모 언어모델(LLM)과 컨텍스트 엔지니어링 기반의 파이프라인을 통해 시의 감정을 2색 배색 팔레트로 시각화하는 함께 읽기(Co-Reading) 시스템을 설계했다. 특히 I.R.I 이미지 스케일을 활용하여 배색의 조화를 고려한 색조 조정 과정을 알고리즘에 적용함으로써, 감정-색채 해석의 일관성과 미적 신뢰도를 확보했다. 또한, 한국어 감정 체계와 색채 반응을 기반으로 구축한 KCoEM을 통해, 서구 중심의 색채 연구에서 충분히 포착하지 못했던 한국적 정동과 색채의 관계를 데이터화했다. 이는 감정 데이터의 문화적 특수성을 색채 언어로 확장하는 시도로서 의의를 지닌다. 이러한 접근은 데이터셋 구축이 하나의 문화적 형식(cultural form)을 구성하는 행위임을 시사한다.

그럼에도 불구하고 본 연구는 몇 가지 분명한 한계를 지닌다. 첫째, KPoEM 데이터셋은 저작권 및 자료 접근의 제약으로 특정 시인군에 한정되어 있어, 문학적 다양성을 충분히 반영하지 못했다. 둘째, KCoEM 데이터셋을 구성하는 색채 데이터의 감정 어휘는 기존 KOTE 44개 감정 범주와 완전한 대응 관계를 형성하지 않으며, 일부 감정 범주에서는 대응 관계가 제한적이다. 또한, KCoEM 데이터셋의 KOTE_id 주석 과정에서 감정과 색의 연관성을 일정 부분 주관적 판단과 언어 기반 분류에 의존하고 있어, 문화적·맥락적 다양성을 충분히 반영하지 못한 한계를 가진다. 따라서 향후에는 KOTE 감정 체계에 대응하는 대규모·표준화된 색채 심리 실험 데이터셋을 구축하고, 이를 기반으로 감정-색채 관계에 대한 문화적·실증적 검증을 수행할 필요가 있다. 셋째, 본 연구에서 제안한 함께 읽기(Co-Reading)는 인공지능을 공동 독자(co-reader)로 이해할 수 있는지에 대한 개념적·존재론적 논의를 여전히 필요로 하며, 개별 독자의 다양한 해석 경험을 완전히 포괄하지 못한다는 한계를 가진

다. 넷째, 본 연구에서 구현한 웹 인터페이스는 해석을 중립적으로 재현하기보다 사용자의 해석 과정과 행동을 일정 부분 유도할 가능성을 내포하고 있다. 따라서 앞으로는 본 연구에서 구현한 읽기 인터페이스를 대상으로 사용성 평가와 사용자 경험 연구를 통해 인간과 인공지능의 상호작용이 문학 해석 과정에 미치는 영향과 인문학적 의의를 실증적으로 검증할 필요가 있다.

본 연구의 의의는 인간의 감정 데이터를 감각적 경험으로 확장 가능한 환경적 미디어로 전환한 점에 있다. 특히 시의 감정을 색채로 시각화함으로써, 독자는 시를 ‘읽는 것’을 넘어 ‘보는 것’으로 경험하게 되며, 이는 문학적 수용의 공감각적 충위를 형성한다. 나아가 본 연구는 데이터셋을 기반으로 인간과 인공지능이 상호작용하는 해석 환경을 구축함으로써, 인공지능을 고정된 결과를 산출하는 도구가 아니라 인간과 함께 의미를 구성하는 존재로 확장했다. 이러한 맥락에서 인공지능은 인간과 함께 문학 텍스트를 읽고 해석하며, 의미를 구성하는 공동 독자(co-reader)로 이해될 수 있다. 이러한 함께 읽기 방법론은 다양한 대규모 언어모델 및 외부 데이터 자원과의 연계를 통해, 시뿐만 아니라 소설·수필 등 다양한 문학 장르와 문화 텍스트로 확장될 수도 있을 것이다.

또한, 본 연구는 이러한 연구 결과를 웹 인터페이스 환경으로 구현함으로써, 데이터셋을 하나의 시각적 매체 환경으로 확장했다. 이는 데이터셋을 미디어로 조망한 본 연구의 이론적 관점이 구체적인 시각 자료 형태로 드러난 결과이기도 하다. 가까이 읽기, 멀리서 읽기, 그리고 함께 읽기의 구조를 통합한 인터페이스는 사용자가 시의 다양한 해석 과정 자체를 경험하도록 설계한 환경이다. 이러한 접근은 데이터 시각화를 정량적 효율성 중심의 정보 전달 수단에서 나아가, 인간과 비인간의 상호작용이 발생하는 해석 공간으로 이해함으로써, 문학 데이터 시각화를 감각적 경험과 해석 행위를 중심으로 하는 상호작용 환경으로 확장할 가능성을 제시한다.

결과적으로 함께 읽기(Co-Reading)는 데이터 시각화를 결과의 표현이 아니라 해석이 발생하는 조건으로 전환하는 시도이며, 인간과 인공지능의 관계를 새롭게 조망하는 하나의 방법론적 가능성을 제안한다. 이러한 의미에서 함께 읽기 시스템은 인간과 인공지능이 데이터셋을 매개로 문학 텍스트를 함께 읽고, 상호작용 속에서 새로운 의미를 구성하는 포스트쓰기체계(postwriting)의 한 형태로 이해될 수 있을 것이다.

인공지능 시대에 인간에게 요구되는 것은 인간과 인공지능이 함께 읽고 해석하며, 나아가 새로운 의미를 구성할 수 있는 환경 자체를 설계하는 일이라고 할 수 있다. 이때 인간은 비로소 인공지능이 생성한 결과를 소비하는 사용자에게 머무르지 않을 수 있다. 따라서 본 연구에서 제안한 함께 읽기(Co-Reading)는 인간과 인공지능이 공존하는 시대에 새로운 해석 환경과 상호작용 구조를 탐색하기 위한 하나의 실험적 모델로 이해될 수 있다. 나아가 본 연구에서 제안한 함께 읽기 시스템과 인터페이스는 향후 연구를 통해 보다 정교화될 수 있으며, 인터페이스를 통한 데이터 시각화 역시 하나의 미학적·매체적 연구 영역으로 확장될 가능성을 지닌다.

【참고문헌】

0. AI 활용 고지

본 논문 초안의 구체화 및 문장 교정, 그리고 데이터 분석 및 시스템 코드 작성에 AI 도구를 활용했으며, 제시된 내용은 연구자의 최종 검토 및 수정을 거쳤다. 본 연구에서는 연구 수행 과정에서 ChatGPT의 GPT-5 및 GPT-5.5 모델을 활용했다.

OpenAI. (2026). *ChatGPT* [Large language model]. <https://chatgpt.com/>

1. 연구 데이터 및 코드

임이로. (2026). *Co-Reading* [Software]. GitHub.

<https://github.com/poet-developer/Co-Reading>

임이로. (2026). *KCoEM* [Data set]. Zenodo.

<https://doi.org/10.5281/zenodo.21131097>

임이로. (2026). *KPoEMInterface* [Software]. GitHub.

<https://github.com/poet-developer/KPoEMInterface>

임이로. (2026). *KPoEMInterface* [Web app]. GitHub Pages.

<https://poet-developer.github.io/KPoEMInterface/>

AKS-DHLAB. (2025a). *KPoEM* [Data set]. Hugging Face.

<https://doi.org/10.57967/hf/6303>

AKS-DHLAB. (2025b). *KPoEM* [Software]. GitHub.

<https://github.com/AKS-DHLAB/KPoEM>

AKS-DHLAB. (2025c). *KPoEM* [Software]. Hugging Face.

<https://doi.org/10.57967/hf/6301>

Lim, I., Ji, H., Koo, S., Jung, S., Yun, J., & Kim, B. (2025). *KPoEM dataset*

[Data set]. Zenodo. <https://doi.org/10.5281/zenodo.15598092>

2. 단행본 및 보고서

국가기술표준원. (2009). *공공디자인 색채표준 가이드*.

<https://www.kats.go.kr/content.do?cmsid=85>

- 김수영. (1974). *거대한 뿌리*. 민음사.
- 백석. (2023). *백석 전 시집: 나와 나타샤와 흰 당나귀*. 스타북스.
- 신석정 전집 간행위원회. (2009). *신석정 전집 1*. 국학자료원.
- 아이알아이. (1998). *색채기호 유형 추출에 관한 연구*. 산업자원부.
- 오규원. (2017). *현대시작법*. 문학과지성사.
- 이정아. (2013). *디자인 미술교육을 위한 색채*. 지식과감성.
- 정현희. (2018). *실제 적용중심의 미술치료* (2판). 학지사.
- 한강. (2013). *서랍에 저녁을 넣어 두었다*. 문학과지성사.
- 한용운. (2016). *님의 침묵 1*. 더스토리.
- Crawford, K. (2021). *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
<https://doi.org/10.2307/j.ctv1ghv45t>
- Day-Lewis, C. (1984). *The poetic image*. J. P. Tarcher. (원저 출판 1947)
- D'Ignazio, C., & Klein, L. F. (2020). *Data feminism*. MIT Press.
- Drucker, J. (2014). *Graphesis: Visual forms of knowledge production*. Harvard University Press.
- Goethe, J. W. von. (2003). *색채론* (권오상, 장희창 옮김). 민음사. (원저 출판 1810)
- Goldsmith, K. (2023). *문예 비창작: 디지털 환경에서 언어 다루기* (길예경, 정주영 옮김). 위크룸프레스. (원저 출판 2011)
- Iser, W. (1980). *The act of reading: A theory of aesthetic response*. Johns Hopkins University Press.
- Manovich, L. (2001). *The language of new media*. MIT Press.
- Manovich, L. (2020). *Cultural analytics*. MIT Press.
- McLuhan, M. (1964). *Understanding media: The extensions of man*. McGraw-Hill.
- Moretti, F. (2013). *Distant reading*. Verso.
- Nijdam, N. A. (2006). *Mapping emotion to color*. University of Twente.
- Peters, J. D. (2015). *The marvelous clouds: Toward a philosophy of elemental media*. University of Chicago Press.

3. 학위 논문

- 김미진. (2023). *색채대비가 점, 선, 면, 형에 미치는 영향* [석사학위논문, 홍익대학교].

<https://doi.org/10.23174/hongik.000000029745.11064.0000322>

김옥. (2017). *감정단어카드를 활용한 유아의 색채반응 연구* [석사학위논문, 광주여자대학교]. <http://www.riss.kr/link?id=T14370612>

왕성필. (2022). *서정주 전집 내 어휘 사용에 대한 정량적 연구 : 키워드, 주제, 감성을 중심으로* [석사학위논문, 한국과학기술원].

<https://koasas.kaist.ac.kr/handle/10203/308294>

4. 연구 논문

강우규. (2024). 국문장편소설 감정 데이터 모델링의 현실적 방안 고찰: 감정분류체계 정립을 중심으로. *민족문학사연구*, 84, 397-432.

<https://doi.org/10.23296/minmun.2024..84.397>

강우규, 김바로. (2024). 국문장편소설의 감정에 대한 정성적·정량적 연구 시론 -〈창선 감의록〉과 〈소현성록〉을 중심으로-. *고소설연구*, 58, 223-258.

<https://doi.org/10.23836/kornov.2024.58.223>

구인모. (2002). 시, 혹은 조선시란 무엇인가-김억의 「작시법(1925)」에 대하여. *한국문학연구*, 25, 297-316.

https://www.kci.go.kr/kciportal/landing/article.kci?arti_id=ART002029401

김병준, 천정환. (2020). 박사학위 논문(2000~2019) 데이터 분석을 통해 본 한국 현대 문학 연구의 변화와 전망. *상허학보*, 60, 443-517.

<https://doi.org/10.22936/sh.60..202010.012>

김애경, 오윤경. (2016). 긍정 및 부정 정서어휘에 대한 색 반응 연구. *조형미/디자인학*, 19(1), 59-66.

<https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE11109635>

김지혜, 박연선. (2015). 2색 배색을 통한 색채 감성 연구. *한국색채학회논문집*, 29(1), 37-48. <https://doi.org/10.17289/jkscs.29.1.201502.37>

김학성. (1997). 한,일 교육용 표색계 비교 연구. *Archives of Design Research*, 819-824. <https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE00879337>

노연숙. (2015). 색의 속성 변화에 따른 이미지의 감성 반응 분석 연구. *한국색채학회 논문집*, 29(2), 49-60. <https://doi.org/10.17289/jkscs.29.2.201505.49>

문재호, 신진우. (2021). 감정을 활용한 직관적 색채 배색 연구: 창의적 색채 디자인 교육 모형 연구. *인문사회과학연구*, 29(1), 47-67.

<https://doi.org/10.22924/jhss.29.1.202105.003>

- 박명건, 박성홍, 신현정. (2022). 등장인물 감성 네트워크에 기반한 소설 작품 군집화. *정보과학회 컴퓨팅의 실제 논문지*, 28(6), 327-331.
<https://doi.org/10.5626/KTCP.2022.28.6.327>
- 박민수. (1986). 김억의 초기시의 상상력과 본질 - 시집 <해파리의 노래>를 중심으로. *관악어문연구*, 11, 113-133. <https://hdl.handle.net/10371/153661>
- 박상민, 나철원, 최민성, 이다희, 온병원. (2018). Bi-LSTM 기반의 한국어 감성사전 구축 방안. *지능정보연구*, 24(4), 219-240.
<https://doi.org/10.13088/jiis.2018.24.4.219>
- 박종빈. (2010). 한국 문학에 표현된 색 이미지의 상징성에 관한 고찰 -현대 시문학을 중심으로-. *조형미디어학*, 13(3), 65-70.
<https://www.kci.go.kr/kciportal/ci/sereArticleSearch/ciSereArtiView.kci?ciSereArticleSearchBean.artId=ART001477479>
- 손선주, 박미숙, 박지은, 손진훈. (2012). 한국어 감정표현단어의 추출과 범주화. *감성과학*, 15(1), 105-120. <https://scienceon.kisti.re.kr/srch/selectPORSrchArticle.do?cn=JAKO201217355622765>
- 신용목. (2019). '詩畫一如'와 '情景의 시학' - 이미지즘과 정지용 시 읽기. *국제언어문학*, 43, 193-220. <https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE08769028>
- 여지혜, 정병범, 함준석, 박준형, 고일주. (2008). 감정표현의 직관성을 향상시키기 위한 색상을 이용한 감정모델의 개발과 적용. *HCI 2008*, 152-157.
<https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE01885251>
- 이문걸. (2002). 한국 현대시의 원형 심상 연구. *어문학*, 78, 397-420.
<https://www.kci.go.kr/kciportal/ci/sereArticleSearch/ciSereArtiView.kci?ciSereArticleSearchBean.artId=ART000851341>
- 이윤진. (2016a). 면적비에 따른 2색배색의 색채조화와 심리메트릭량의 관계. *한국색채학회논문집*, 30(1), 127-136.
<https://doi.org/10.17289/jkscs.30.1.201602.127>
- 이윤진. (2016b). 색채조화도 및 선호도에 대한 평가구조의 비교연구 - 배색특징을 중심으로 -. *한국색채학회논문집*, 30(2), 121-130.
<https://doi.org/10.17289/jkscs.30.2.201605.121>
- 장세은, 박호민, Suo, M., 이성화. (2021). 감정분석 시각화를 통한 소설과 영화의 비교 : 모비딕 중심으로. *언어과학*, 28(1), 207-229.
<https://doi.org/10.14384/kals.2021.28.1.207>
- 전성규. (2024). 이광수의 오감(五感) 오케스트레이션 - 소리 데이터 생산과 사운드 맵

- (sound map)을 중심으로. *춘원연구학보*, 29, 73-109.
<https://doi.org/10.31809/crj.2024.12.31.073>
- 정민영, 이현수. (2019). 딥러닝에 기반을 둔 배색이미지의 감성어휘 추출 방법론. *한국실내디자인학회 학술대회논문집*, 112-115.
<https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE09274992>
- 최석재. (2008). 감정동사의 유형과 그 의미특성. *어문논집*, 58, 127-159.
https://www.kci.go.kr/kciportal/landing/article.kci?arti_id=ART001290487
- 한지운, 나건. (2021). 한국 성인의 긍정과 부정 감정 색채 특징. *한국디자인리서치*, 6(3), 322-335. <https://doi.org/10.46248/kidrs.2021.3.322>
- 황정석, 조택연. (2024). 텍스트의 감정 분석 기반 조형 시각화 방법 연구 - VAD 감정 차원과 조형 언어 사전을 중심으로 -. *한국공간디자인학회 논문집*, 19(5), 11-24.
<https://doi.org/10.35216/kisd.2024.19.5.11>
- Abdul-Rahman, A., Lein, J., Coles, K., Maguire, E., Meyer, M., Wynne, M., Johnson, C. R., Trefethen, A., & Chen, M. (2013). Rule-based visual mappings - With a case study on poetry visualization. *Computer Graphics Forum*, 32(3pt4), 381-390. <https://doi.org/10.1111/cgf.12125>
- Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 2623-2631). Association for Computing Machinery. <https://doi.org/10.1145/3292500.3330701>
- Anantrasirichai, N., & Bull, D. (2021). Artificial intelligence in the creative industries: A review. *Artificial Intelligence Review*, 55, 589-656.
<https://doi.org/10.1007/s10462-021-10039-7>
- Anderson, M., Motta, R., Chandrasekar, S., & Stokes, M. (1996). Proposal for a standard default color space for the Internet-sRGB. *Proceedings of the IS&T 4th Color Imaging Conference* (pp. 238-245).
<https://doi.org/10.2352/CIC.1996.4.1.art00061>
- Demszky, D., Movshovitz-Attias, D., Ko, J., Cowen, A., Nemade, G., & Ravi, S. (2020). GoEmotions: A dataset of fine-grained emotions [Preprint]. *arXiv*. <https://doi.org/10.48550/arXiv.2005.00547>
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 248

- 255). <https://doi.org/10.1109/CVPR.2009.5206848>
- Haider, T., Eger, S., Kim, E., Klinger, R., & Menninghaus, W. (2021). PO-EMO: Conceptualization, annotation, and modeling of aesthetic emotions in German and English poetry [Preprint]. *arXiv*.
<https://doi.org/10.48550/arXiv.2003.07723>
- Ismayilzada, M., Paul, D., Bosselut, A., & van der Plas, L. (2025). Creativity in AI: Progresses and challenges [Preprint]. *arXiv*.
<https://doi.org/10.48550/arXiv.2410.17218>
- Jänicke, S., Franzini, G., Cheema, M. F., & Scheuermann, G. (2015). On close and distant reading in digital humanities: A survey and future challenges. *EuroVis 2015*. <https://doi.org/10.2312/eurovisstar.20151113>
- Jeon, D., Lee, J., & Kim, C. (2024). User guide for KOTE: Korean online that-gul emotions dataset. *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)* (pp. 17254–17270).
<https://aclanthology.org/2024.lrec-main.1499/>
- Lim, I., Ji, H., & Kim, B. (2026). KPoEM: A human-annotated dataset for emotion classification and RAG-based poetry generation in Korean modern poetry. *The Review of Korean Studies*, 29(1), 161–206.
<https://doi.org/10.25024/review.2026.29.1.006>
- Liu, C., Osama, M., & de Andrade, A. (2019). DENS: A dataset for multi-class emotion analysis [Preprint]. *arXiv*.
<https://doi.org/10.48550/arXiv.1910.11769>
- McCurdy, N., Lein, J., Coles, K., & Meyer, M. (2016). Poemage: Visualizing the sonic topology of a poem. *IEEE Transactions on Visualization and Computer Graphics*, 22(1), 439–448.
<https://doi.org/10.1109/TVCG.2015.2467811>
- Mei, L., Yao, J., Ge, Y., Wang, Y., Bi, B., Cai, Y., Liu, J., Li, M., Li, Z.-Z., Zhang, D., Zhou, C., Mao, J., Xia, T., Guo, J., & Liu, S. (2025). A survey of context engineering for large language models [Preprint]. *arXiv*.
<https://doi.org/10.48550/arXiv.2507.13334>
- Öhman, E., Bizzoni, Y., Feldkamp Moreira, P., & Nielbo, K. (2024). EmotionArcs: Emotion arcs for 9,000 literary texts. *Proceedings of the 8th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heri*

- tage, Social Sciences, Humanities and Literature (LaTeCH-CLfL 2024)* (pp. 51-66). <https://doi.org/10.18653/v1/2024.latechclfl-1.7>
- Ou, L.-C., Luo, M. R., Woodcock, A., & Wright, A. (2004a). A study of colour emotion and colour preference. Part I: Colour emotions for single colours. *Color Research & Application*, 29(3), 232-240. <https://doi.org/10.1002/col.20010>
- Ou, L.-C., Luo, M. R., Woodcock, A., & Wright, A. (2004b). A study of colour emotion and colour preference. Part II: Colour emotions for two-colour combination. *Color Research & Application*, 29(4), 292-298. <https://doi.org/10.1002/col.20024>
- Picca, D., & Pavlopoulos, J. (2024). Deciphering emotional landscapes in the Iliad: A novel French-annotated dataset for emotion recognition. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)* (pp. 4462-4467). <https://aclanthology.org/2024.lrec-main.399/>
- Reagan, A. J., Mitchell, L., Kiley, D., Danforth, C. M., & Dodds, P. S. (2016). The emotional arcs of stories are dominated by six basic shapes. *EPJ Data Science*, 5, Article 31. <https://doi.org/10.1140/epjds/s13688-016-0093-1>
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161-1178. <https://doi.org/10.1037/h0077714>
- Sreeja, P. S., & Mahalakshmi, G. S. (2019). PERC-An emotion recognition corpus for cognitive poems. In *2019 International Conference on Communication and Signal Processing (ICCSP)* (pp. 200-207). <https://doi.org/10.1109/ICCSP.2019.8698020>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *Neural Information Processing Systems*. (pp. 5998-6008). <https://mlanthology.org/neurips/2017/vaswani2017neurips-attention/>
- Wan, Q., Hu, S., Zhang, Y., Wang, P., Wen, B., Lu, Z., ... Liu, S. (2024). "It felt like having a second mind": Investigating human-AI co-creativity in prewriting with large language models. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW1), Article 84, 1-26. <https://doi.org/10.1145/3637361>

Wang, B., Shi, Q., Wang, X., Zhou, Y., Zeng, W., & Wang, Z. (2025). Emotion Lens: Interactive visual exploration of the circumplex emotion space in literary works via affective word clouds. *Visual Informatics*, 9(1), 84-98. <https://doi.org/10.1016/j.visinf.2025.02.003>

5. 웹 자원

국가기술표준원. (n.d.). *색채표준정보*. <https://www.kats.go.kr/content.do?cmsid=83>

국립국어원. (n.d.). *한국어-영어 학습사전 (Korean-English Learners' Dictionary)*. <https://krdict.korean.go.kr/eng>

김소월. (n.d.). *5일밤 산보*. 위키문헌. https://ko.wikisource.org/wiki/5일밤_산보

디아리움. (2020). *IRI색채연구소 - 색채 이미지 공간*. <https://www.engedi.kr/library/?bmode=view&idx=4236116>

위키백과. (n.d.). *배색*. <https://ko.wikipedia.org/wiki/배색>

위키백과. (n.d.). *보색*. <https://ko.wikipedia.org/wiki/보색>

위키백과. (n.d.). *파인튜닝*. https://ko.wikipedia.org/wiki/파인_튜닝

위키백과. (n.d.). *환각 (인공지능)*. [https://ko.wikipedia.org/wiki/환각_\(인공지능\)](https://ko.wikipedia.org/wiki/환각_(인공지능))

위키백과. (n.d.). *Analogous colors*. https://en.wikipedia.org/wiki/Analogous_colors

이상. (n.d.). *꽃나무*. 위키문헌. <https://ko.wikisource.org/wiki/꽃나무>

이상. (n.d.). *오감도*. 위키문헌. <https://ko.wikisource.org/wiki/오감도>

임이로. (2025). *데이터셋의 환경적 전환: 인공지능과 함께 읽기(Co-Reading)를 통한 한국 근현대 시-감정-색채 멀티모달 미디어* [Conference presentation]. 2025 중앙대학교 영화미디어학센터 대학원생 컨퍼런스. Zenodo. <https://doi.org/10.5281/zenodo.17850832>

서울신문. (2007, October 15). *시인협회 '한국 10대 시인' 선정*. <https://www.seoul.co.kr/news/society/2007/10/15/20071015027028>

정지용. (n.d.). *향수*. 위키문헌. <https://ko.wikisource.org/wiki/향수/향수>

한용운. (n.d.). *당신을 보았습니다*. 위키문헌. https://ko.wikisource.org/wiki/님의_침묵/당신을_보았습니다

한용운. (n.d.). *타골의 시를 읽고*. 위키문헌. [https://ko.wikisource.org/wiki/님의_침묵/타골의_시\(GARDENISTO\)를_읽고](https://ko.wikisource.org/wiki/님의_침묵/타골의_시(GARDENISTO)를_읽고)

Google. (2026). *Gemini 2.5 Flash* [Large language model]. Gemini API.

<https://ai.google.dev/gemini-api/docs/models/gemini-2.5-flash>

Hart, M. (1971). *Project Gutenberg*. <https://www.gutenberg.org/>

Hitokun. (n.d.). *PCCS tone map* [Web app]. GitHub Pages.
<https://hitokun-s.github.io/old/tool/pccs-tone-map.html>

Jeon, D. (2024). *KOTE* [Data set]. GitHub. <https://github.com/searle-j/KOTE>

Kim, B., & Kang, W. (2025). *KENT (Kang woo kyu Emotion for Novel Text)* [Data set]. OSF. <https://doi.org/10.17605/OSF.IO/XSY8U>

Kisel, N., Volkov, I., Hanzelkova, K., Janouskova, K., & Matas, J. (2025). *Flaws of ImageNet, computer vision's favorite dataset*. ICLR Blogposts 2025. <https://iclr-blogposts.github.io/2025/blog/imagenet-flaws/>

Lee, J. (2022). *KcELECTRA* [Software]. GitHub.
<https://github.com/Beomi/KcELECTRA>

Lim, I., & Kim, B. (2026). *Co-Reading: A human-AI co-reader media system for poetry-emotion-color* [Poster presentation]. Hong Kong Association for Digital Humanities (HKADH), The Chinese University of Hong Kong (CUHK). Zenodo. <https://doi.org/10.5281/zenodo.18753718>

Mueller, A. (2026). *wordcloud* (v1.9.6) [Software]. GitHub.
https://github.com/amueller/word_cloud

Optuna. (2025). *Optuna* (v4.6.0) [Software]. GitHub.
<https://github.com/optuna/optuna>

Park, J. (2021). *GoEmotions-Korean* [Data set]. GitHub.
<https://github.com/monologg/GoEmotions-Korean>

Plotly. (2026). *Plotly.py* (v6.8.0) [Software]. GitHub.
<https://github.com/plotly/plotly.py>

Puchner, M. (n.d.). *Custom GPTs and online education*.
<https://www.martinpuchner.com/custom-gpts-and-online-education.html>

Python Software Foundation. (2026). *colorsys – Conversions between color systems*. In Python 3 documentation.
<https://docs.python.org/3/library/colorsyst.html>

Python Software Foundation. (2026). *Python* [Software].
<https://www.python.org/>

R Core Team. (2025). *R: A language and environment for statistical computing* (v4.5.2) [Software]. R Foundation for Statistical Computing.
<https://www.R-project.org/>

- Saito, H. (2013). *PCCS color code data* [Data set]. GitHub Gist.
<https://gist.github.com/hidex7777/5268678>
- SHAP. (2026). *SHAP* (v0.51.0) [Software]. GitHub.
<https://github.com/shap/shap>
- The pandas development team. (2025). *pandas-dev/pandas: Pandas* (v2.3).
Zenodo. <https://doi.org/10.5281/zenodo.17229934>
- Thornhill, J. (2024, February 1). *Why AI hallucinations can be a good thing*.
Financial Times.
<https://www.ft.com/content/3b88cbd7-e72d-48c7-badc-096006488c36>
- Wickham, C. (2024). *munsell: Utilities for using Munsell colours* (v0.5.1)
[R package]. <https://CRAN.R-project.org/package=munsell>
- Wikimedia Foundation. (2024). *Kowikisource dump files* [Data set].
<https://dumps.wikimedia.org/kowikisource/>

Abstract

Co-Reading with AI: Color Visualization of Emotion Data in Korean Modern Poetry¹⁾

I RO LIM

Major in Cultural Informatics

The Graduate School of Korean Studies

The Academy of Korean Studies

This study proposes a “*Co-Reading*” system as an environment in which humans and artificial intelligence collaboratively interpret literary texts. To achieve this goal, this study reconceptualizes datasets as environmental media that mediate interpretation between humans and AI. It also redefines AI as a “*co-reader*” that participates in the construction of meaning alongside humans. To implement such an interpretive environment, this study focuses on *poetry*, a literary genre characterized by interpretive ambiguity and multiplicity. Furthermore, recognizing *color* as an expressive element that mediates emotions and imagery in poetry, this study designed and developed a human–AI Co-Reading system that visualizes the emotions embedded in poetry through color.

1) This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2025S1A5B5A20019820).

This thesis is partly based on a collaborative research project co-authored by the author involving the construction of the KPoEM dataset and the development of the KPoEM emotion classification model. The related work is published as follows:

Lim, I., Ji, H., & Kim, B. (2026). KPoEM: A human-annotated dataset for emotion classification and RAG-based poetry generation in Korean modern poetry. *The Review of Korean Studies*, 29(1), 161–206.
<https://doi.org/10.25024/review.2026.29.1.006>

The KCoEM dataset, the emotion-to-color transformation algorithm, the Co-Reading framework, and the web interface presented in this thesis were independently developed by the author.

First, this study constructed KPoEM (Korean Poetry Emotion Mapping), an emotion dataset that structures human interpretations of emotion in Korean modern poetry, and developed a KPoEM emotion classification model for poetic texts.

Second, based on existing studies in color psychology, this study constructed KCoEM (Korean Color Emotion Mapping), a Korean color-emotion dataset, and developed an emotion-to-color transformation algorithm that incorporates emotional attributes and principles of color harmony.

Third, this study designed a Co-Reading system that reconstructs human emotional responses and AI interpretations of an input poetic text into a two-color palette through context engineering. To realize this process, the communication structure and pipeline of the Co-Reading system were designed and implemented.

Fourth, using the KPoEM dataset, this study realized *close reading*, *distant reading*, and *co-reading* based on human-AI interaction as three independent reading environments and integrated them into a web-based interface prototype for comparative exploration. Whereas close reading and distant reading focus on providing microscopic and macroscopic perspectives on poetic texts, respectively, co-reading aims to encourage human sensory reinterpretation through color images reconstructed by humans and AI. Through this approach, this study suggests that data visualization can function as a communicative environment in which humans and AI collaboratively interpret texts and construct meaning.

Ultimately, this study experimentally implements a digital environment in which humans and AI collaboratively read the same literary text and construct new meanings through datasets and interfaces. Furthermore, it proposes one possible direction for a *postwriting* in an era where artificial intelligence increasingly mediates human knowledge and creativity.

Keywords: Co-Reading, Human-AI Interaction, Environmental Media, Postwriting, Emotion Dataset, Emotion-Color Transformation, Multimodality, Data Visualization, Web Interface, Affective Computing, Korean Modern Poetry, Digital Humanities

【부록】

[부록 1] KOTE 감정 어휘와 KCoEM 감정 어휘 대응 관계¹⁾

KOTE ID	KOTE 감정 어휘	감정군	KOTE 영문 표기	클러스터 예시 어휘	영어 뜻풀이	KCoEM 감정 어휘
0	불평/불만	분노	complaint /dissatisfaction	dissatisfied, oppose, criticize, complaint	dissatisfaction	불만
1	환영/호의	-	welcome/favor	welcome, approval, kindness, enthusiastic	enthusiastic	enthusiasm
2	감동/감탄	기쁨	impressed /admiration	admiring, great, praise, compliment	admiration	감동
3	지긋지긋	분노	fed up	fed up, struggle, arduous, sick and tired	sick, tired	sick, tired
4	고마움	기쁨	gratitude	praiseworthy, commendable, favor, blessing, mercy	gratitude, blessing	감사하다, 축복
5	슬픔	슬픔	sadness	sad, miss, lonely, tear	sadness, sad, lonely	슬픔, 외롭다, 외로움, sadness, sad, lonely
6	화남/분노	분노	anger/rage	anger, rage, obsessed, fury	anger, rage	분노, 화, anger, rage
7	존경	-	respect	respect, loyal, veneration, follow, obedience	admiration	감동
8	기대감	-	anticipation	new, achieve, together, harmonious, vitality	expectation	기대하다, 기대감, 기대

9	우쭐댐/무시함	기쁨/분노	arrogance /disregard	arrogance, pompous, ignore, bragging, boast, gasconade	arrogance	arrogant
10	안타까움/실망	슬픔	pitifulness /disappointment	disappointed, sorry, upset, deplorable, regretful	deplorable	애통, 답답하다, 낙담, 후회
11	비장함	슬픔	resolute	resolute, determination	determination	determination
12	의심/불신	두려움	doubt/distrust	suspicious, doubtful, lie	distrust	distrust
13	뿌듯함	기쁨	pride	successful, victory, worthwhile, accomplish	pride, self-respect	자존, pride
14	편안/쾌적	기쁨	comfort/cozy	comfortable, ease, cozy, cool, warm	comfort	편안하다, 위로, warm, comfort, comfortable
15	신기함/관심	-	curiosity/interest	interested, curious	interesting	재미있다
16	아껴주는	-	caring	caring, adore, dear	care, love	사랑하다, 사랑
17	부끄러움	슬픔	shame	ashamed, humiliated	shame	부끄러움, shame
18	공포/무서움	두려움	fear/scary	fear, anxious, tense, pressed	fear	공포, 두려움, fear, fearful
19	절망	-	despair	frustrated, joys & sorrows, hurt, grief, letdown	despair, frustrated	절망, frustration
20	한심함	-	pathetic	pathetic, belittled, stupid, impudence	stupid	답답하다
21	역겨움/징그러움	분노	disgust /repulsiveness	disgusted, repulsive, dirty	disgust	disgust

22	짜증	분노	irritation	irritated, pissed off, ridiculous	irritation, irritated	irritating
23	어이없음	슬픔	preposterous	dumbfounded, stunned, stuffy, enervated, WTF	stuffy	답답하다
24	없음	-	NO EMOTION	-	-	no-emotion, no/emotion
25	패배/자기혐오	-	defeat/self-hatred	failure, miserably, extorted	failure	패배, 실패
26	귀찮음	분노	laziness	bothered, dawdling	laziness	게으름
27	힘듦/지침	-	fatigue /exhaustion	tired, peak, exhausted	tired	tired
28	즐거움/신남	기쁨	pleasure /excitement	excited, funny	excited	즐거움, 신남, excited
29	깨달음	-	realization	realize, enlightened, wakened, conviction, belief	belief	faith
30	죄책감	-	guilt	guilt, blamed, repentance, remorse	guilt	죄책감
31	증오/혐오	분노	loathing/hatred	loathing, hatred, scorn, vilifying	scorn	경멸, 증오
32	흐뭇함 (귀여움/예쁨)	기쁨 (사랑)	pleased (cute/pretty)	handsome, pretty, sweet, thrilled, cute, aegyo	satisfied, cute, sweet	만족하다, 만족, 귀여움, 달콤함, pleasant, satisfaction
33	당황/난처	두려움	embarrassment	embarrassed, disconcerted,	embarrassed	embarrassed

		/슬픔	/awkwardness	awkward, untoward		
34	경악	놀람	shock	shocked, flabbergasted, pass out, freaked out	-	놀라움
35	부담/안_내킴	-	burden /unwillingness	unwilling, denial, pressure, cannot be bothered, give up	denial, pressure	거부당함, 중압감
36	서러움	슬픔	sorrow	sorrowful, mirthless, weary, sobbing, upset, complicated	sorrow	sorrow
37	재미없음	-	boredom	bored, tedium, trite, dull	bored	따분함, bored
38	불쌍함/연민	슬픔	pity/compassion	pity, sadly, choked up, heartrending	-	슬픔, 외롭다, 외로움, sadness, sad, lonely
39	놀람	놀람	surprise	astonished, startled	surprise	놀라움
40	행복	기쁨	happiness	happy, affection, valuable, hope, luck	happy, hope, luck, happiness	기쁘다, 애정, 행복, 희망, 행운, 희열, happiness, happy
41	불안/걱정	두려움	anxiety/worry	apprehensive, worry, threatened	anxiety, worry	걱정하다, 근심, 불안
42	기쁨	기쁨	joy	delight, ecstasy, love	joy, love	기쁨, 애정, 사랑하다, joy, 사랑, 희열, love
43	안심/신뢰	기쁨	relief/trust	relief, trust, intimate, close	relief	신뢰감, trust

1) KOTE 논문(Jeon et al., 2024)에서는 ‘어이없음’ 클러스터의 예시 어휘를 ‘stuffy’로 표기했으나, 해당 어휘가 영어 사전에 존재하지 않고 문맥상 ‘stuffy’의 오기로 판단되어 본 연구에서는 이를 수정하여 사용했다. 또한 ‘감정군’의 해석은 최석재(2008)의 연구를 참고했다.