

머신러닝 기반 시각화를 통한 악성 댓글 문제 완화 연구

기술적 조치를 통한 혐오 표현 대응

악성댓글 처리 알고리즘을 활용한 댓글 시각화

Profile

박지현 | Jihyun Park

현 랜덤웍스 테크 디렉터

아트애펬크놀로지학과 석사 졸업

컴퓨터공학부 학사 졸업

UC Irvine DRG(Dutt Research Group) Internship 연구원

INDEX

1. Introduction

서론

연구 배경 및 필요성

연구 목적

2. References

관련 연구 및 사례

악성 댓글 문제점 개선 사례

시각화 연구 사례

3. Solution

댓글 시각화 시스템

SG-Comment

뉴스 댓글 시스템

4. Evaluation

실험 및 평가

실험 방법

실험 결과

5. Discussion

논의

6. Conclusion

결론

INDEX

1. Introduction

서론

연구 배경 및 필요성

연구 목적

2. References

관련 연구 및 사례

악성 댓글 문제점 개선 사례

시각화 연구 사례

3. Solution

댓글 시각화 시스템

SG-Comment

뉴스 댓글 시스템

4. Evaluation

실험 및 평가

실험 방법

실험 결과

5. Discussion

논의

6. Conclusion

결론

연구 배경

댓글의 정의

댓글이란?

인터넷에 올린 원문에 대하여 짧게 답하여 올리는 글

이윤정, 조환규, 우균, 지정훈, “**TRIB**: 블로그 댓글 분류 및 시각화 시스템”
정보처리학회논문지 **D16, no. 5 (2009): 817-824**

연구 배경

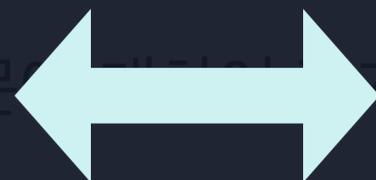
댓글의 정의

긍정적

부정적

자유로운 토론의 장

댓글이란?



다른 사람 헐뜯기
허위사실 퍼뜨리기

Introduction

이찬성, 조원기, 유진, 서정원, "TRIB: 웹로그 댓글 분류 및 시각화 시스템"
정보처리학회논문지 D16, no. 5 (2009): 817-824

“[기고]청소년들의 ‘악플 방지 지도’ 어떻게 하나?,” 불만닷컴, 2015년 3월 31일 수정, 2019년 5월 21일 접속,
<http://www.bulmanzero.com/news/articleView.html?idxno=14337>

연구 배경

악성댓글의 정의

악성댓글이란?

온라인에서 상대방에게 정신적 상처를 줄 수 있는 글을 작성하는 행위

Ybarra, M. and Mitchell, K. Online aggressor/targets, aggressors, and targets: a comparison of associated youth characteristics. *The Journal of Child Psychology and Psychiatry*, 45(7), pp.1308-1316, 2004.

연구 배경

악성댓글 활성화 요인

익명성	비대면성	집단성
온라인 상에서 신분을 노출하지 않는다	상대방과 대면하지 않는다	같은 악의적 내용을 쓰는 사람들이 많을 수록 심리적 부담감은 분산되고 줄어든다

김창환, 악성 댓글의 실태와 대응 방안, 주간기술동향 통권 1437호 2010.3.17

연구 필요성

악성댓글 이슈(Issue)

commented: 내 인생 아끼워서 고소 안하는거지 너봐주는 거 아니다 병신아 5m

다 내놔 또라이야 법정

사람 잡는 '키보드 킬러' 판친다

▶법원도 피싱사이트가 죽을 직전까지 '악성 댓글(악플)'로 괴롭혔다는 사실이 알려지면서 악플이란 말-유해 정보 신고 접수는 2003년 7만3511건이던 해서 8만2161건(2004년)→9만8713건(2005년)→12만9572건(2006년)→19만1488건(2007년)으로 해마다 늘고 있다.

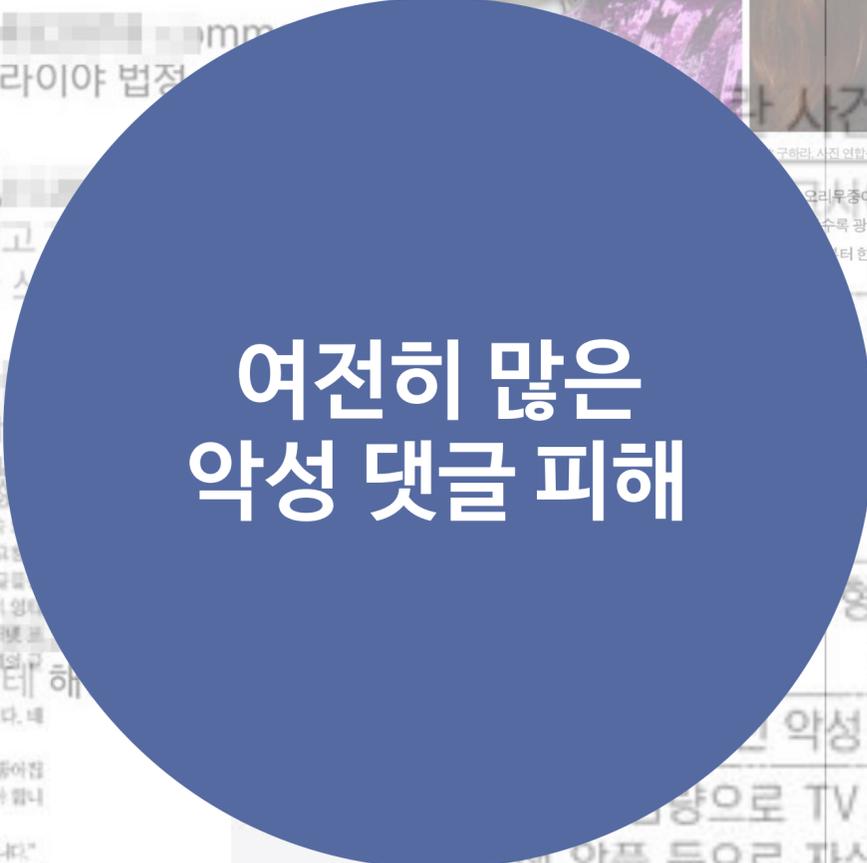
▶이 중 '권리 침해'를 당했다며 신고한 건수는 2005년 806건에서 지난해엔 1만5000건을 넘은 것으로 나타났다. 5년 새 18배 이상 증가한 것이다. 권리 침해란 댓글을 포함해 ▶특정인을 비방하거나 ▶취급 사실을 유포하는 경우 ▶ 명예훼손에 해당하는 글을 게재한 경우 등을 말한다. 분서 결과 권리침해 중 가장 빈도가 높은 건 명예훼손과 허위 사실 적어냈다. 심의기관에 "문제가 있다"고 판단한 건수도 2배에 늘고 있다. 심의기관이 '시정 요구'를 내린 건수는 2005년 4만2000여 건이었으나 지난해에는 11만2200여 건을 기록했다. 이 가운데 '권리 침해'의 문제가 있다며 시정을 요구한 경우도 2005년 1515건에서 지난해 8254건으로 늘었다. 특히 8254건 중 99%에 달하는 문제의 글을 삭제하라고 요구했다. 그만큼 사안이 중대하다고 판단한 것이다.

▶대체별로 보면 지난해의 경우 포털 네이버가 신고-시정요구 건수에서 모두 1위를 기록했으나 올해는 다음이 그 자리를 차지했다.

▶'사이버 명예훼손'도 급증-사람적 차원에서 사이버 범죄, 특히 명예훼손과 관련된 문제도 갈수록 심각해지고 있는 것으로 분석됐다.

▶한나라당 정갑승 의원은 경찰청으로부터 받은 자료에 따르면 '사이버 범죄'는 2004년 7만7099건에서 지난해 8만8847건으로 증가했다. 이 가운데 통신 사기나 개인 정보 유포, 불법 복제 판매 등은 증가 또는 감소폭을 반복하는 반면 명예훼손 등으로 인한 범죄는 상승곡선을 보이고 있다.

▶명예훼손 범죄는 2004년 3667건에서 2007년 1만282건으로 173% 증가했다. 올 들어 상반기까지는 5145건을 기록했다.



다이어트 성공 TV출연 여고생 '인터넷 악플' 시달리다 자살

TV 오락 프로그램에 다이어트 성공사례로 출연했던 한 여고생이 인터넷 악플로 인해 자살했다.

이양은 지난달 2일 SBS TV 프로그램 '놀라운 대회 스타킹'에 출연, 자신의 다이어트 경험을 소개했다. 돌아간 후 할머니를 위해 3개월 만에 87kg에서 47kg으로 몸무게를 줄였다고 밝힌 이양은 "51세 건강화 초거 환자입니다. 내가 자꾸 불려와 거절합니다." "답변" "간이 안 좋으면 체중도 안 떨어집니다. 음식 잘 먹고 허벅지를 마사지해야 합니다. 연락하세요." "질문" "면역력에 대해고민하는 직장인입니다." "답변" "OO제품을 먹으면 신기하게 몸이 가벼워집니다." "영희의 경우엔 이사는 "인터넷에 민원란 잘못된 의학 정보는 명예훼손 못지않은 엄청난 피해를 가져다줄 수 있다"며 대책 마련이 시급하다고 지적했다.

이양 어머니는 경찰조사에서 "전날 새벽 다이어트를 한다면서 밥을 먹지 않아 심한 꾸중을 들은 뒤 곧장 방으로 들어갔다"고 말했다. 이어 "딸이 2005년 7월부터 하루에 우유 한두 컵만 마시는 등 다이어트를 하오며 신경이 날카로워져 있는 상황에서 개인 홈페이지에 악플이 잇따라 올라와 마음고생을 많이 했다"고 덧붙였다.

이양의 방에서는 '어머니에게 꾸지람을 들어 마음이 아프다. 살 뺐으니 저승 가서 할머니를...'이라고 적힌 유서가 발견됐다. 친구 박모양(16)은 "약 먹고 살 뺐다는 악플이 돌면서 고민이 많았다. 지난해엔 인터넷 사이트에 '제가 요즘 악플에 시달리고 있습니다. 제가 어떻게 해야 하나요. 죽기를 바라나요. 정말 힘듭니다. 저도 사람입니다'란 글을 올린 적이 있다"고 전했다.

경찰 관계자는 "이양 부모가 악플을 단 사람들의 처벌을 원하고 수사를 의뢰할 경우 수사를 벌이겠다"고 말했다. 대전 | 윤희일기자

2008년 9월 안재환, 아내 정선희씨 관련 의혹 등으로 자살

2008년 10월 최진실 안재환 관련 사제, 복머 사이버 수사대 사건 의뢰 후 자살

연구 필요성

악성댓글 이슈(Issue)

사람들은 '키보드질러' 판친다

- 연구 목적 -

“댓글 문화를 개선할 수 있는 하나의 해결 방안으로써 시각화, 인터랙션 디자인, 머신 러닝 기술을 활용하여 악성 댓글의 작성과 노출이 모두 자연스럽게 줄어들도록 유도하는 뉴스 댓글 시각화 시스템을 제안”

다이어트 성공 TV 출연 여고생
'인터넷 악플' 시달리다 자살

Introduction

INDEX

1. Introduction

서론

연구 배경 및 필요성

연구 목적

2. References

관련 연구 및 사례

악성 댓글 문제점 개선 사례

시각화 연구 사례

3. Solution

댓글 시각화 시스템

SG-Comment

뉴스 댓글 시스템

4. Evaluation

실험 및 평가

실험 방법

실험 결과

5. Discussion

논의

6. Conclusion

결론

사례1. 악성 댓글 문제점 개선 사례

인터넷실명제

이용자의 실명과 주민등록번호가 확인될 때만 인터넷에 글을 올릴 수 있도록 하는 제도로써 인터넷 상 익명성이 존재하는 환경에서는 여론 왜곡이 불가피하며 공론의 장에서의 자유로운 발언에는 책임감이 따라야한다는 관점에서 대안으로 제시됨



2012년, 헌법이 보장하는 개인 표현 자유를 침해할 뿐만 아니라 공익의 효과도 미미하다는 점을 근거로 위헌 판결을 받아 **폐지**

헌재 2012.08.23 2010헌마47

사례1. 악성 댓글 문제점 개선 사례



사용자와 함께 만드는 댓글 문화

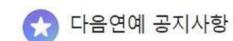
- 1
내가 보고 싶지 않은
댓글이 있다면 바로 **접기요청**하여
해당 댓글을 접을 수 있어요!
- 2
접힌 댓글을 보고 싶을 경우 **펼쳐서**
그 내용을 확인할 수 있어요!
- 3
다수의 요청으로 자동접힌 댓글에 대해
추가 평가도 가능합니다 :)

1
mega****
너희들이 하는 짓이 뭐 그렇지...꼴도 보기 싫은 녀석들
한대 맞아야 정신
2017-05-25 16:3 | 접기 요청
답글 32 | 1221 | 30

2
mega****
Ⓜ 사용자 요청으로 접힌 댓글입니다
2017-05-23 11:02 | 내용 보기

3
mega****
너희들이 하는 짓이 뭐 그렇지...꼴도 보기 싫은 녀석들
한대 맞아야 정신차리지.
2017-05-23 11:02 | 닫기
답글 32 | 1221 | 30
다수가 접기 요청한 댓글입니다.
위 댓글에 대한 의견은? | 접기유지 | 펼침요청

네이버, '접기요청'



연예 뉴스 댓글 폐지 안내

13,053 읽음 · 2019.10.30

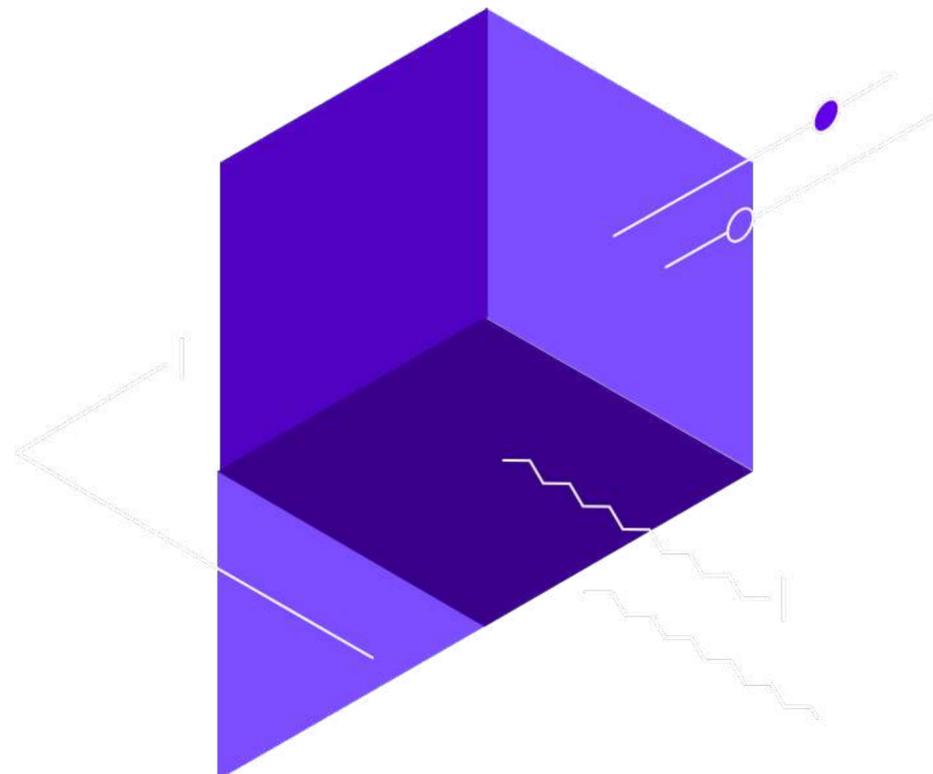
연예뉴스 댓글 서비스를 잠정 폐지합니다

카카오, '연예뉴스 댓글 서비스 폐지'

사례1. 악성 댓글 문제점 개선 사례



네이버, '클린봇'



구글 & 직소, Perspective API

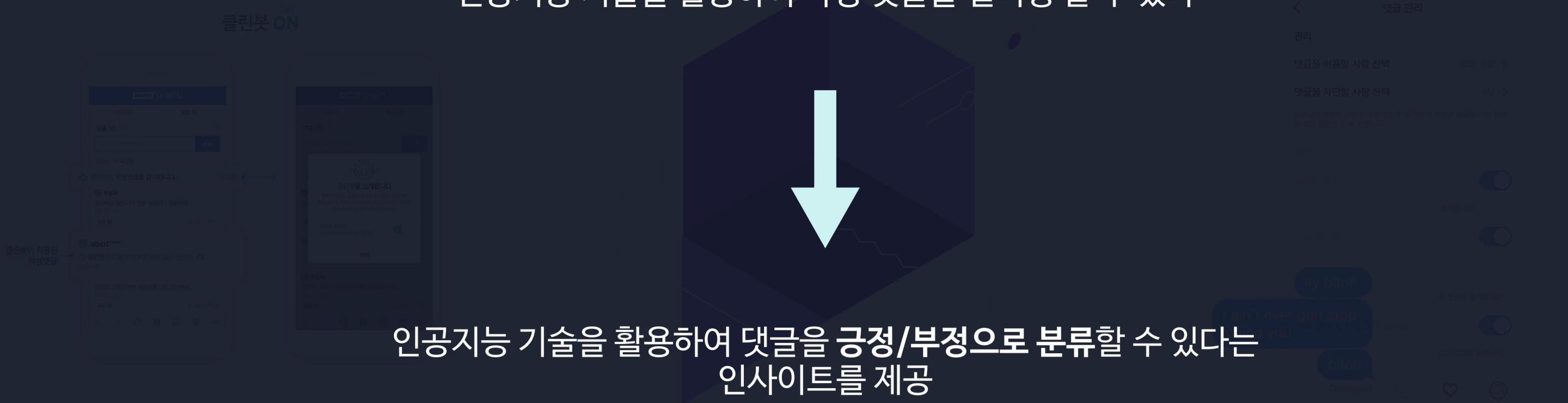


페이스북 & 인스타그램, DeepText

References

사례1. 악성 댓글 문제점 개선 사례

“인공지능 기술을 활용하여 악성 댓글을 필터링 할 수 있다”



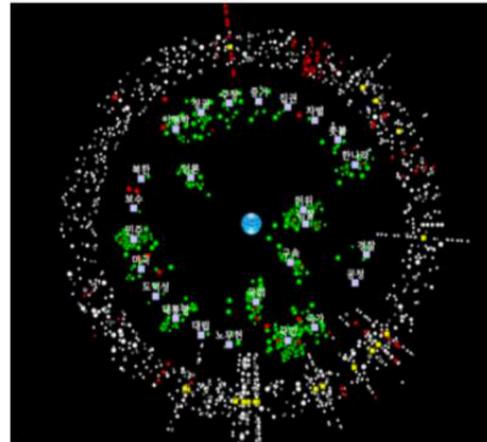
References

네이버, '클린봇'

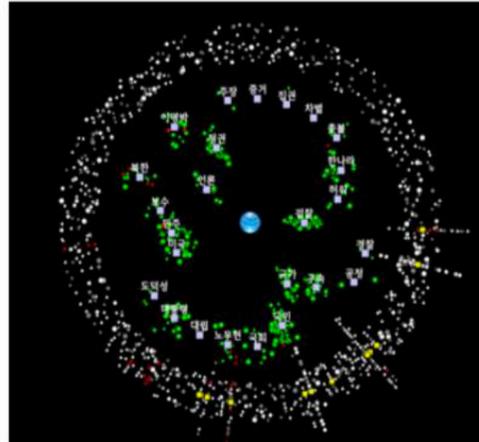
구글&직소, Perspective API

페이스북&인스타그램, DeepText

사례2. 댓글 시각화 연구 사례



(a) S_p 의 시각화



(b) S_S 의 시각화

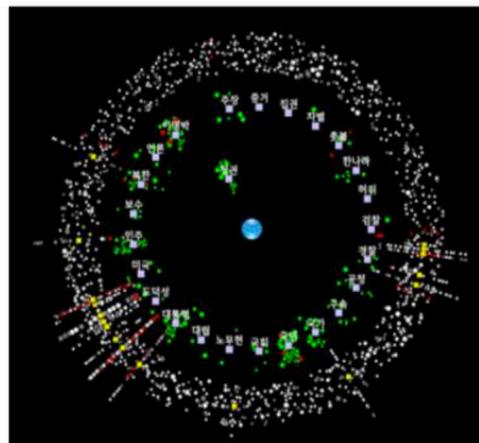
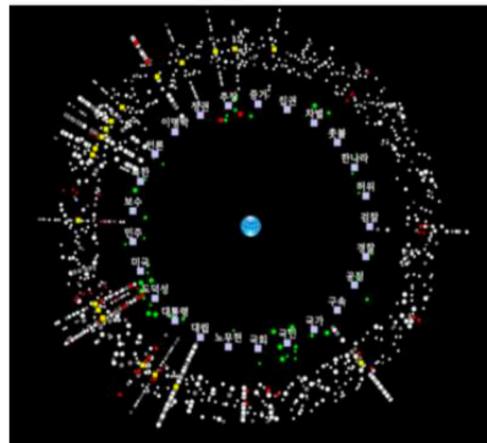


표 4. 시각화 표현 요소

시각화 요소	표현 대상
■	키워드
●	단어로 분류된 댓글
○	단어로 분류되지 않은 댓글
●	스팸 댓글
●	무의미 댓글
○	마우스로 선택한 댓글
○	선택 댓글의 이전 댓글
○	선택 댓글의 이후 댓글
■	10개 이상의 댓글 작성자 id

이윤정, 지정훈, 우균, 조환규. (2009). 인터넷 게시물의 댓글 분석 및 시각화. 한국콘텐츠학회논문지, 9(7), 45-56.

사례2. 댓글 시각화 연구 사례

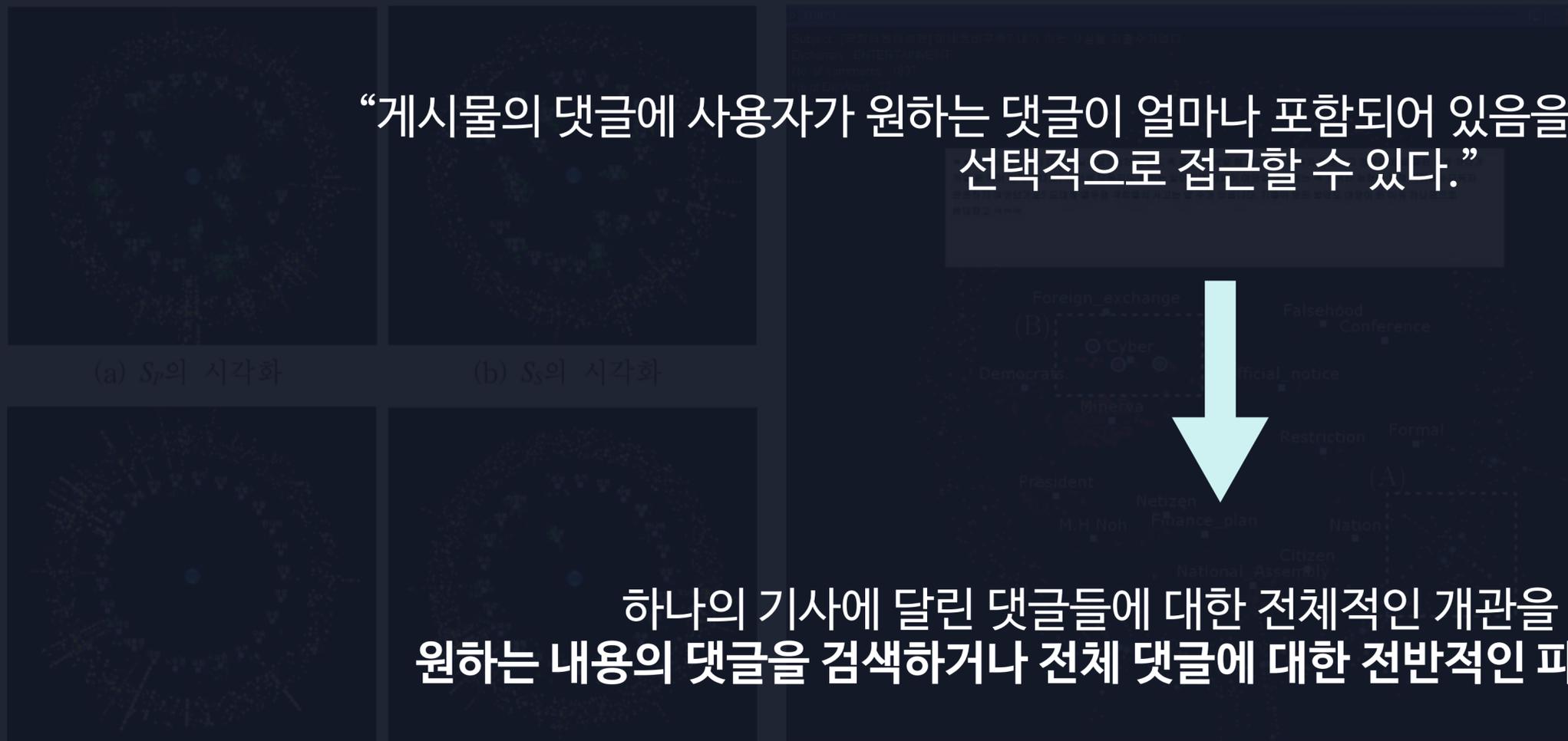


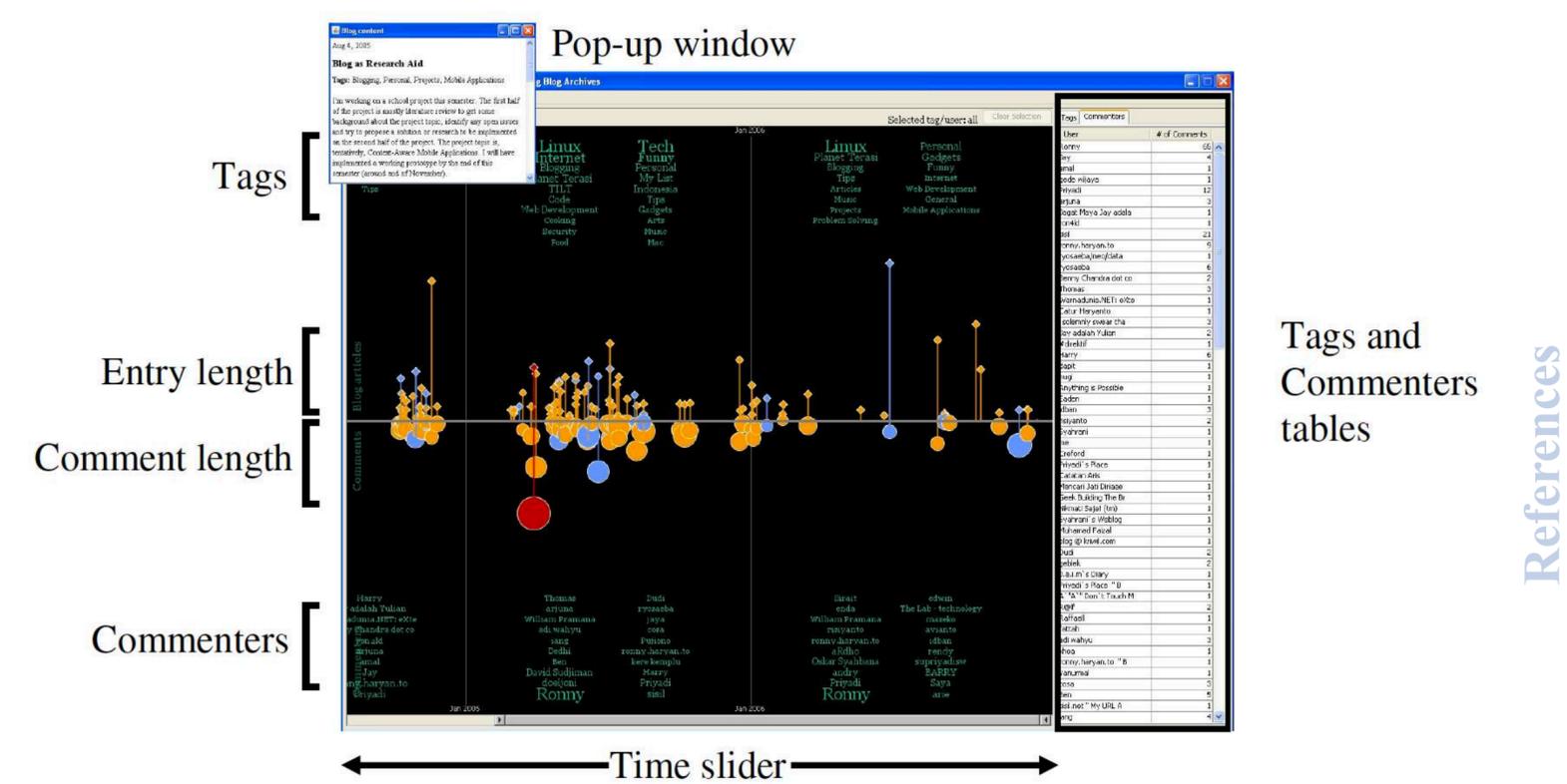
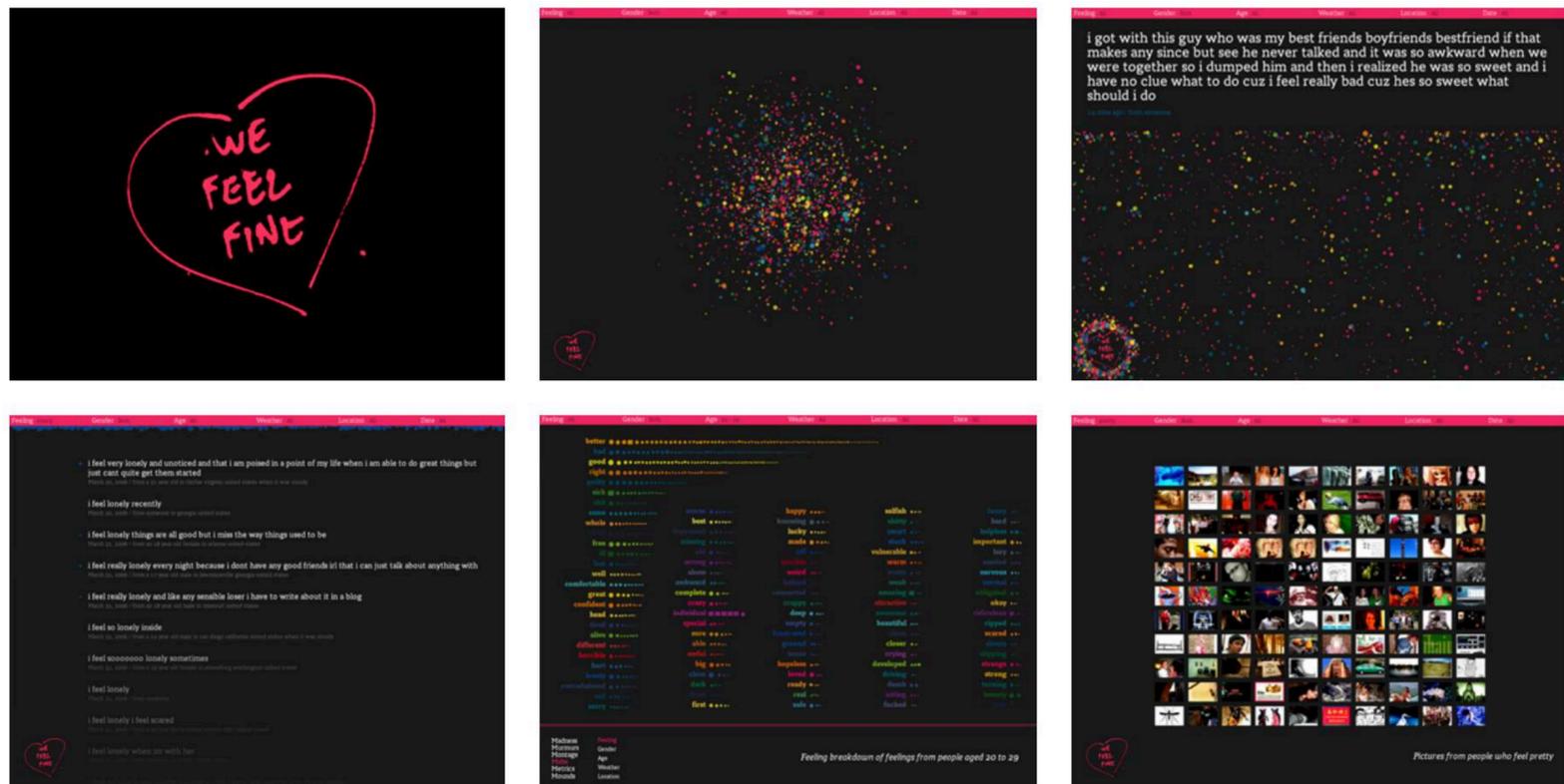
표 4. 시각화 표현 요소

시각화 요소	표현 대상
□	키워드
●	단어로 분류된 댓글
○	단어로 분류되지 않은 댓글
●	스팸 댓글
○	무의미 댓글
●	마우스로 선택한 댓글
●	선택 댓글의 이전 댓글
●	선택 댓글의 이후 댓글
■	10개 이상의 댓글 작성자 id

References

이윤정, 지정훈, 우균, 조환규. (2009). 인터넷 게시물의 댓글 분석 및 시각화. 한국콘텐츠학회논문지, 9(7), 45-56.

사례3. 블로그 시각화 연구 사례



Jonathan Harris, We Feel Fine, 2006, [online] Available at: www.wefeelfine.org [Accessed 11 Mar. 2020]

Indratmo, Julita Vassileva, and Carl Gutwin, "Exploring blog archives with interactive visualization," In Proceedings of the Working Conference on Advanced Visual Interfaces, pp.:39-46, 2008

사례4. 댓글 분석 및 분류 연구 사례

기준	선행 연구	특징
댓글의 내용 분석	G. Mishne, N. Glance, 2006	댓글의 논쟁 정도 예상
스팸 댓글 분류	Mishne, G., Carmel, D., Lempel, R., 2005	블로그의 본문과 댓글, 댓글이 링크된 페이지 간의 유사도 비교를 통해 스팸 여부 판단
	L. Xiao-bing, N. Zhang, 2006	클러스터링을 이용하여 스팸 필터링
	W. -F. Hsiao, T. -M. Chang, G. -H. Hu, 2007	
악성 댓글 분류	배민영, 이윤정, 지정훈, 우균, 조환규, 2009	악성 댓글의 분류 성능 개선

사례4. 댓글 분석 및 분류 연구 사례

기준	선행 연구	특징
댓글의 내용 분석	G. Mishne, N. Glance, 2006	댓글의 논쟁 정도 예상
악성 댓글 분류	Mishne, G., Carmel, D., Lempel, R., 2005 사용자가 보다 자율적인 판단으로 올바른 댓글 문화를 스스로 만들어 갈 수 있도록 유도하는 시각화 시스템을 제안한다는 점에서 차별점을 지닌다.	블로그의 본문과 댓글, 댓글이 링크된 페이지 간의 유사도 비교를 통해 스팸 여부 판단
	W. -F. Hsiao, T. -M. Chang, G. -H. Hu, 2007	클러스터링을 이용하여 스팸 필터링
악성 댓글 분류	배민영, 이윤정, 지정훈, 우균, 조환규, 2009	악성 댓글의 분류 성능 개선

INDEX

1. Introduction

서론

연구 배경 및 필요성

연구 목적

2. References

관련 연구 및 사례

악성 댓글 문제점 개선 사례

시각화 연구 사례

3. Solution

댓글 시각화 시스템

SG-Comment

뉴스 댓글 시스템

4. Evaluation

실험 및 평가

실험 방법

실험 결과

5. Discussion

논의

6. Conclusion

결론

SG-Comment 뉴스 댓글 시스템

4가지 특징

Feature.1	Feature.2	Feature.3	Feature.4
댓글의 긍정/부정 시각화	색상 및 시간 지연 방식의 악성댓글 필터링	한 화면 내 전체 댓글 접근 기능	댓글 중심의 뉴스 기사 시각화

특징 1. 댓글의 긍정/부정 시각화

4가지 특징

Feature.1

Feature.2 - 2가지 사용 목적 -

Feature.3

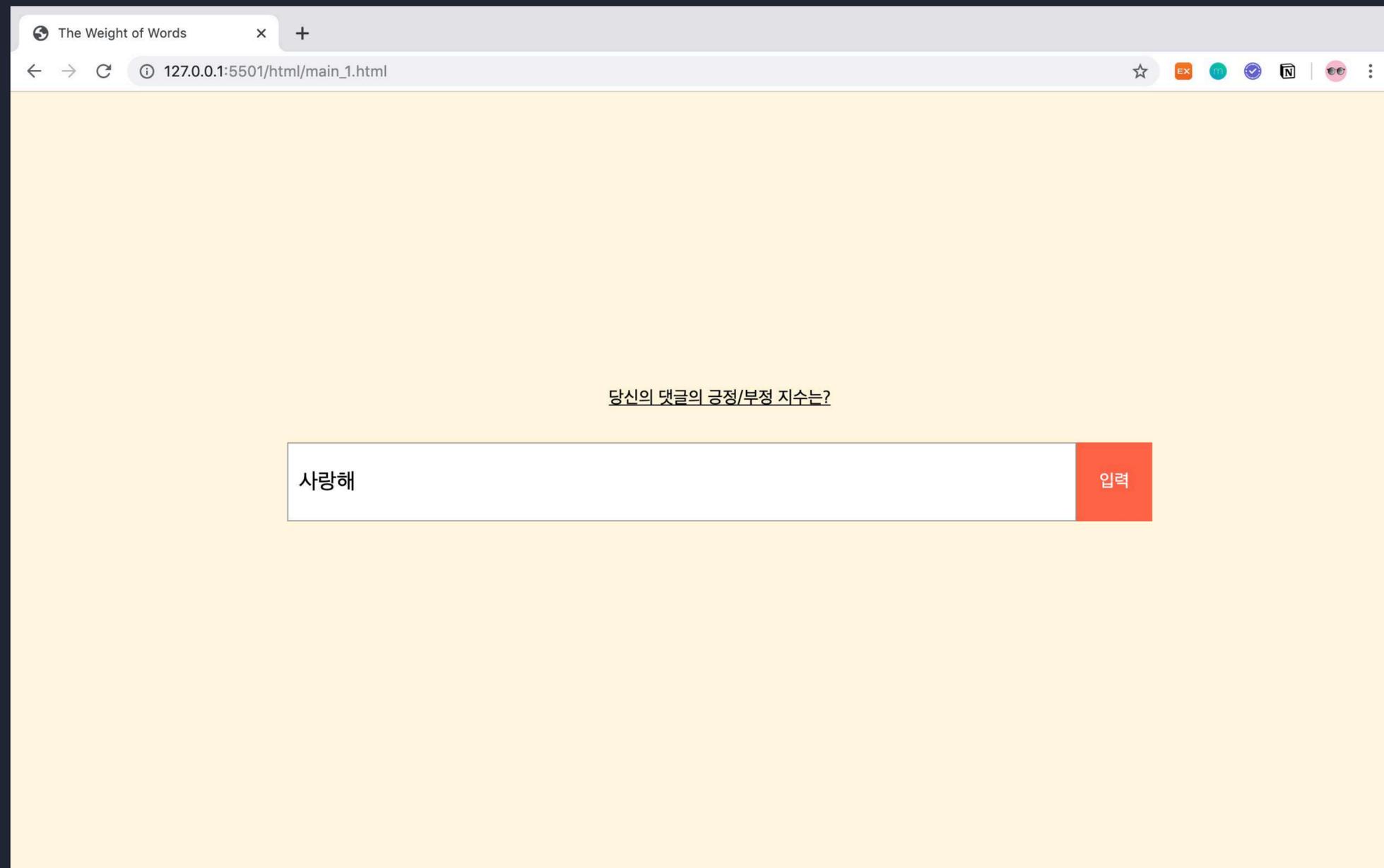
Feature.4

- 1. 사용자가 입력하고 있는 댓글의 긍정, 부정 정도를 실시간으로 판단할 수 있다.
- 2. 전체 댓글의 긍정, 부정 경향을 파악하고, 그에 따른 댓글 검색 기능을 제공한다.

Solution

특징 1. 댓글의 긍정/부정 시각화

4가지 특징



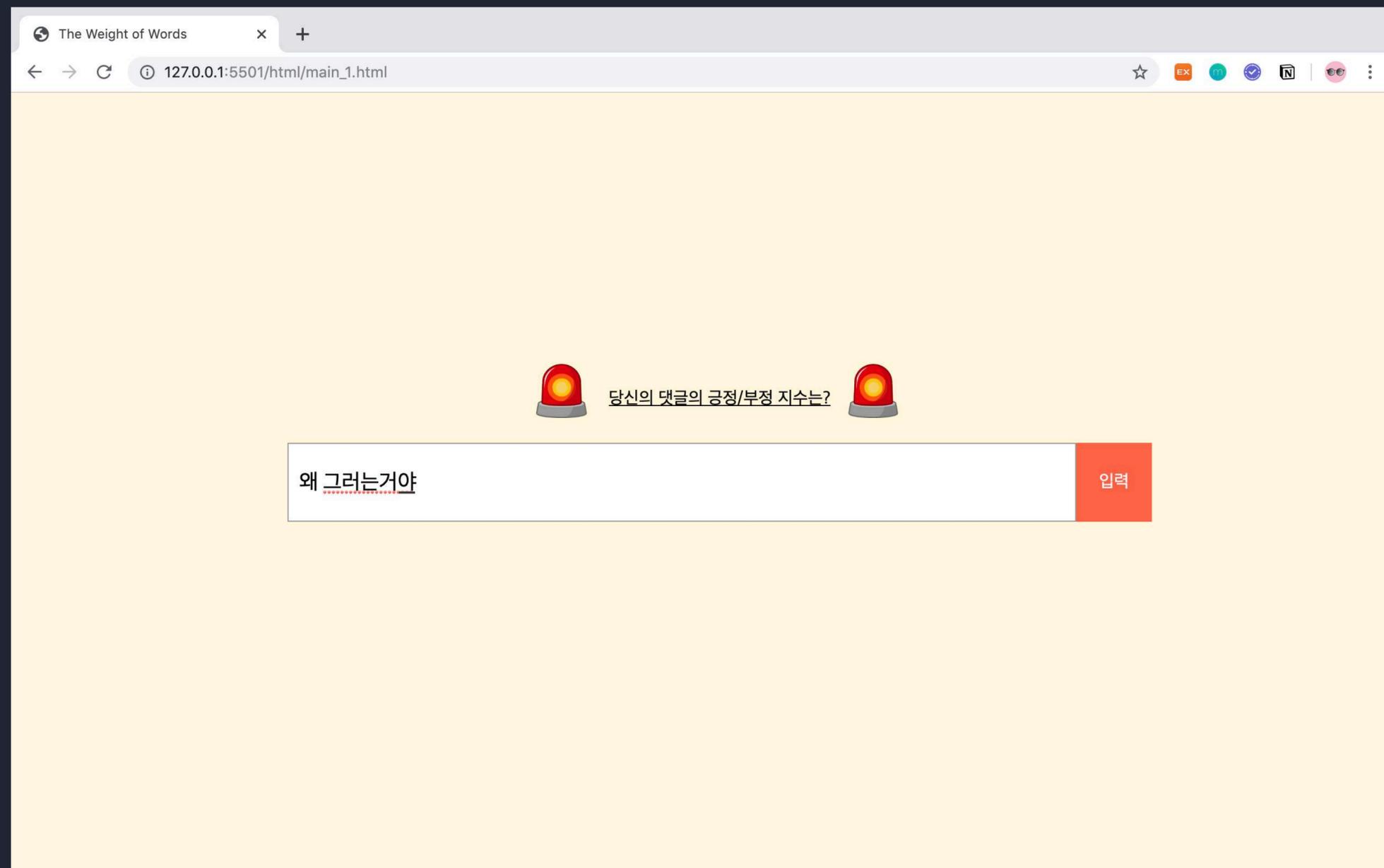
| 긍정적인 댓글 작성

어떠한 아이콘도 나타나지 않는다.

Solution

특징 1. 댓글의 긍정/부정 시각화

4가지 특징

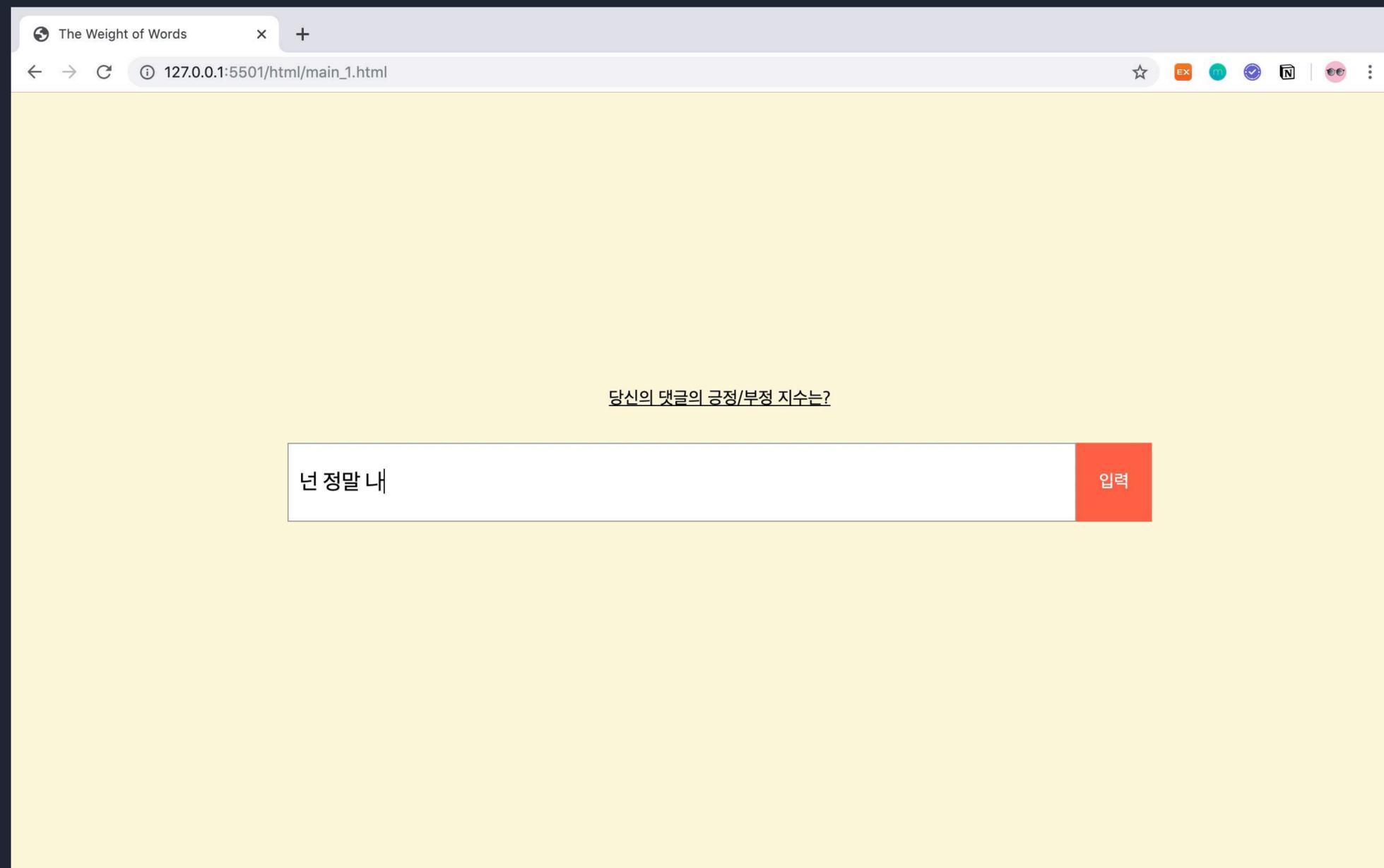


| 부정적인 요소가
50% 이상인 댓글

사이렌 아이콘 이미지를 나타내어
악성 댓글이 될 수도 있음을
1차적으로 경고

특징 1. 댓글의 긍정/부정 시각화

4가지 특징



| 부정적인 요소가
70% 이상인 댓글

사이렌이 울리는 애니메이션이
활성화되어 보다 강한 시각적 경고
제공

Solution

특징 1. 댓글의 긍정/부정 시각화

4가지 특징

Feature.1

Feature.2 - 2가지 사용 목적 -

Feature.3

Feature.4

- 1. 댓글의 긍정/부정 사용자가 입력하고 있는 댓글의 긍정, 부정 정도를 실시간으로 판단할 수 있다.
- 2. 전체 댓글의 긍정, 부정 경향을 파악하고, 그에 따른 댓글 검색 기능을 제공한다.

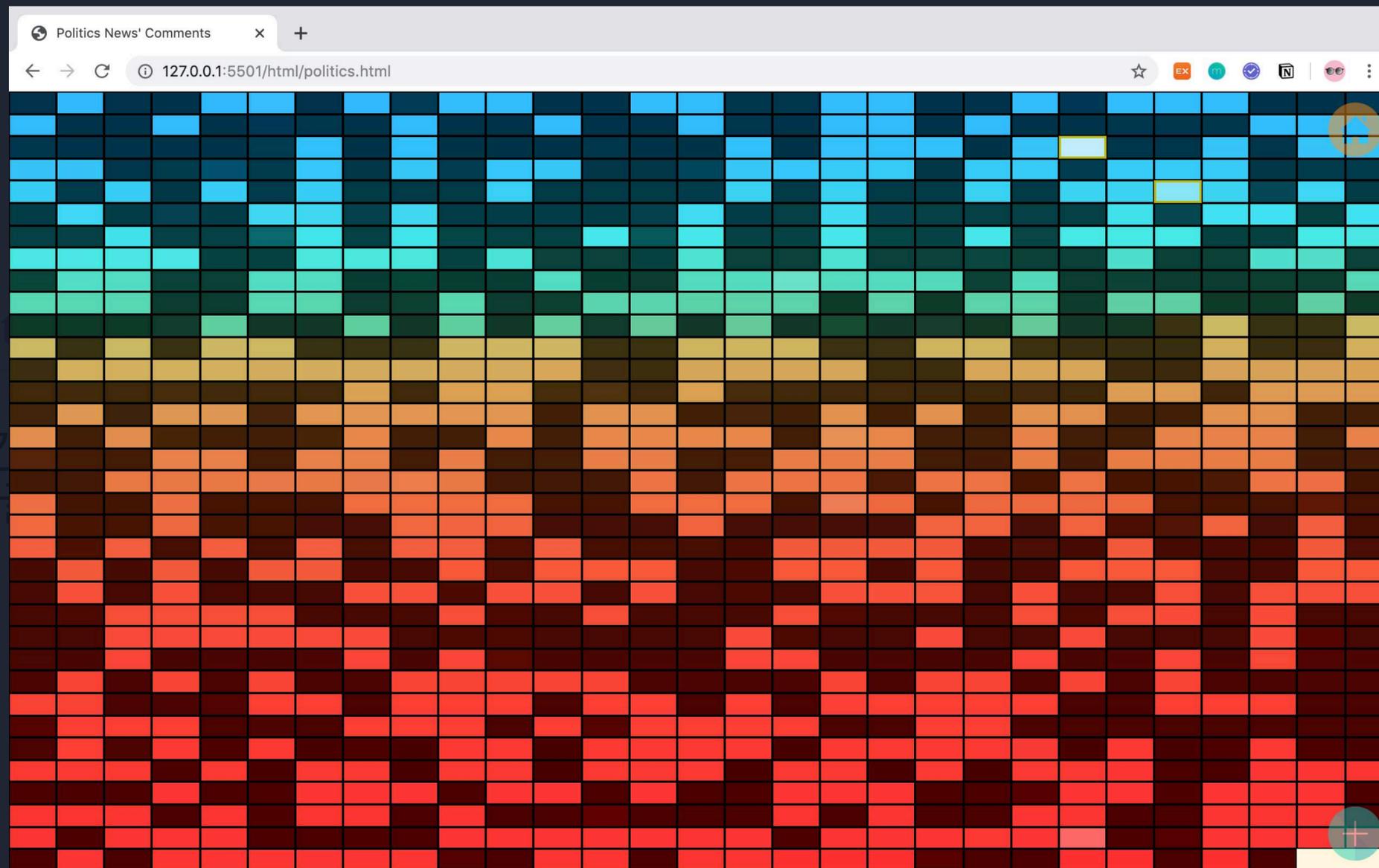
Solution

특징 1. 댓글의 긍정/부정 시각화

4가지 특징

색조 및 채도를 이용한 댓글의 긍/부정도 시각화
 (Visualizing positiveness/negativeness of a comment using hue and saturation)

	Comment Negativeness: 100 -> 0	Comment Positiveness: 0 -> 100
Hue (0 ~ 359)	0(red) -> 100(green)	100(green) -> 200(blue)
Saturation (0% ~ 100%)	100% -> 50%	50% -> 100%



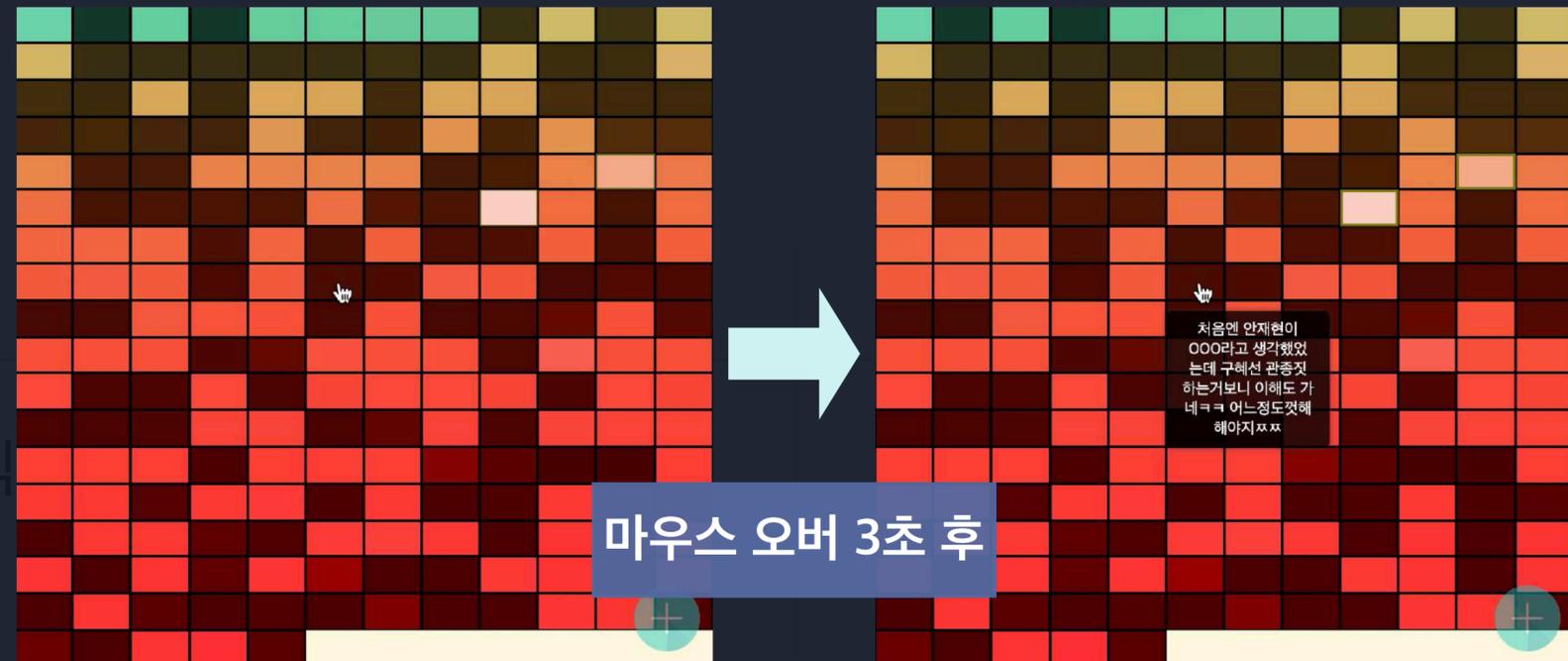
Solution

특징 2. 색상과 시간 지연 기반 악성 댓글 필터링

4가지 특징

2-단계 댓글 내용 노출 필터링
(2-level comment exposure filtering design)

	Negative comments having more dislikes than likes	All other comments
Level-1 Filtering	Colored rectangles with low brightness	Colored rectangles with normal brightness
Level-2 Filtering	Time delay (3 seconds)	-

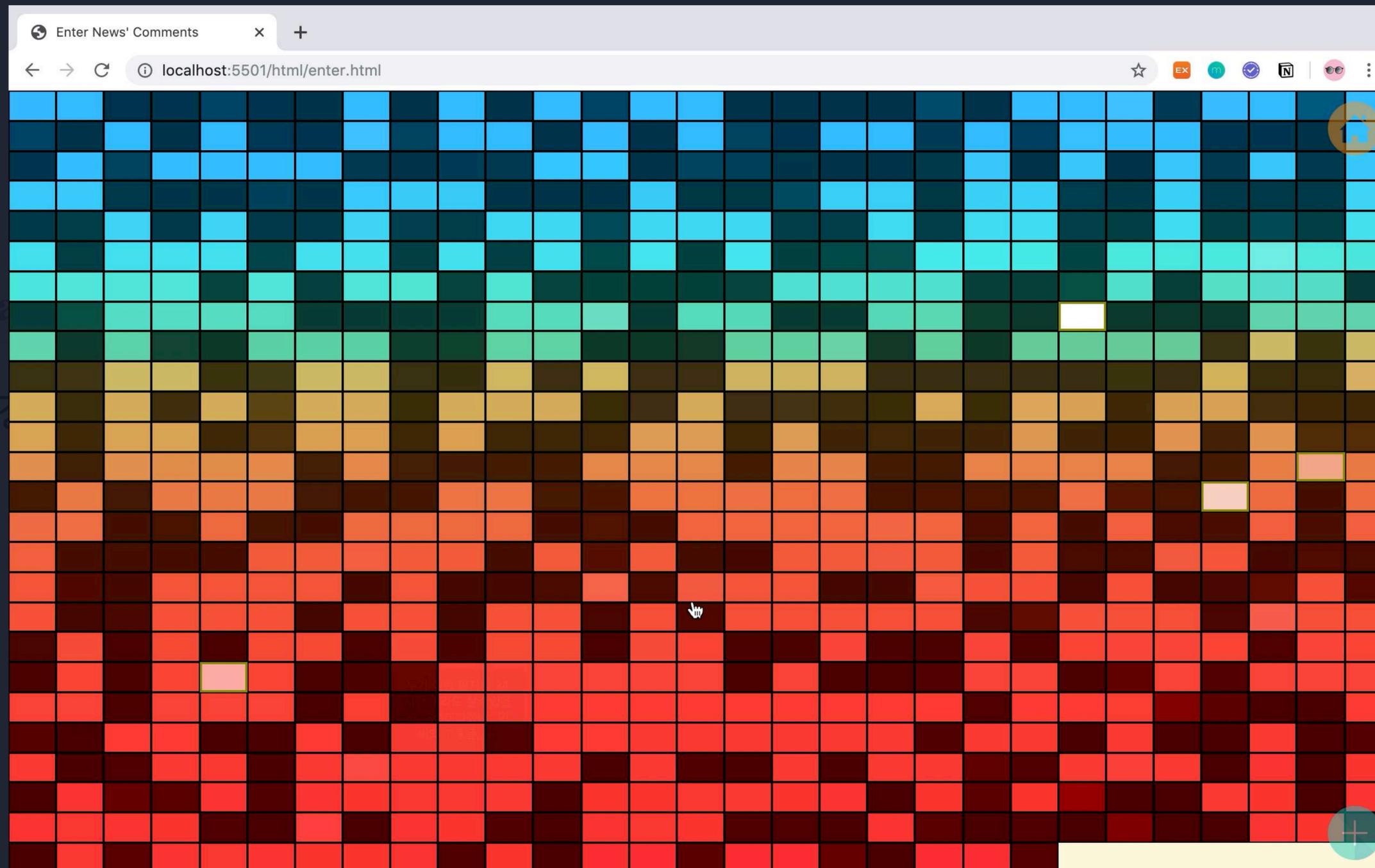


Solution

부정도가 높은 댓글(부정도 70% 초과) 중 다른 사용자들에 의해 비공감(dislikes)을 더 많이 받은 경우 해당 사각형은 낮은 명도(50%)로 표현되며
마우스 포인터를 해당 사각형에 위치해도 바로 내용이 보여지지 않도록 하여 노출도를 제한한다.

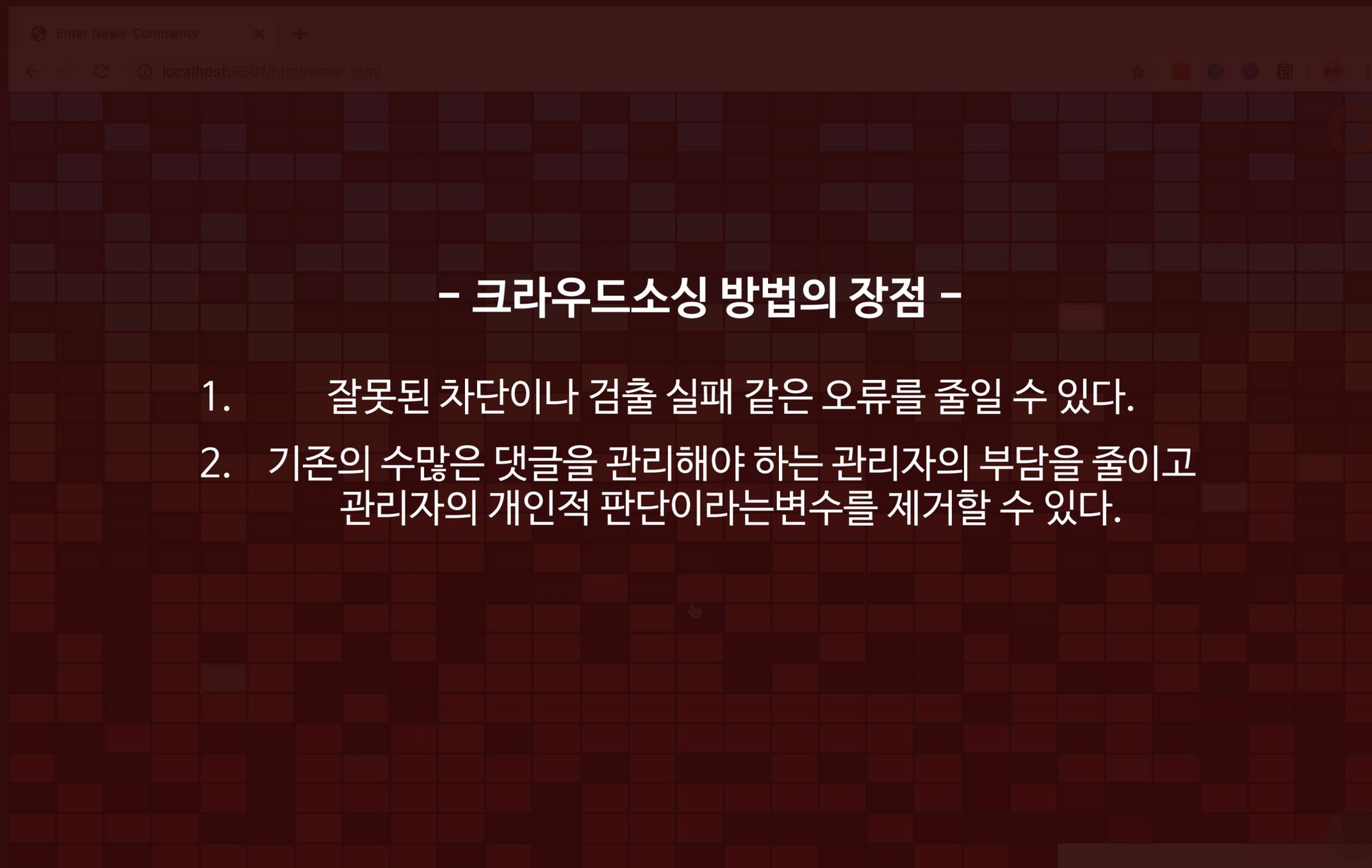
특징 2. 색상과 시간 지연 기반 악성 댓글 필터링

4가지 특징



Solution

특징 2. 색상과 시간 지연 기반 악성 댓글 필터링



특징 3. 한 화면 내 전체 댓글 접근 가능

4가지 특징

Feature.1

ACOP (All Comments on One Page)

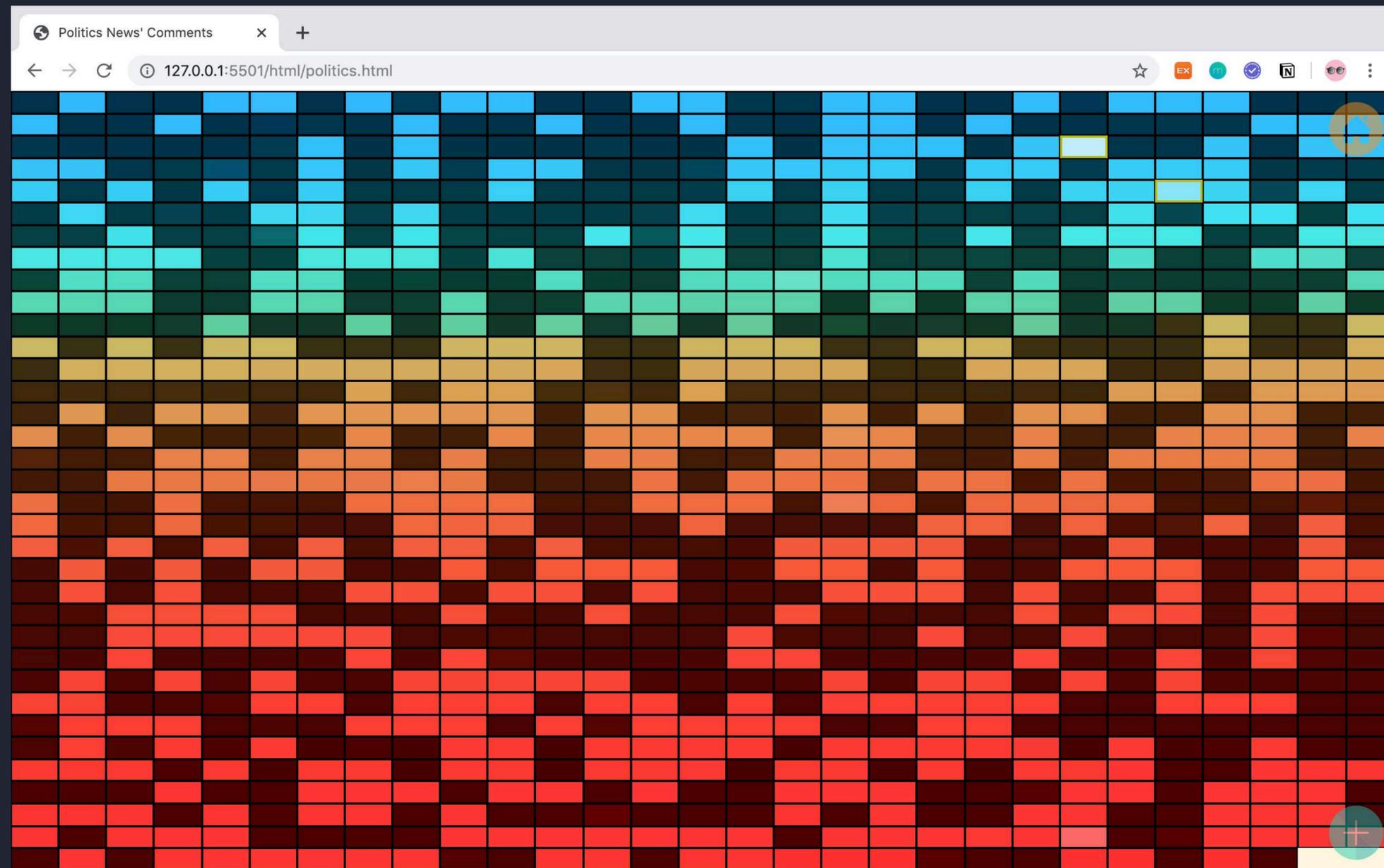
Feature.4

댓글 기존의 뉴스 플랫폼에서 사용하는 복수의 페이지 상에서의 댓글 열거 (listing)와 화면 스크롤 방식으로 댓글을 보여주는 방법에서 탈피하여 한 페이지에서 모든 댓글을 접근할 수 있도록 디자인한 방식

Solution

특징 3. 한 화면 내 전체 댓글 접근 가능

4가지 특징

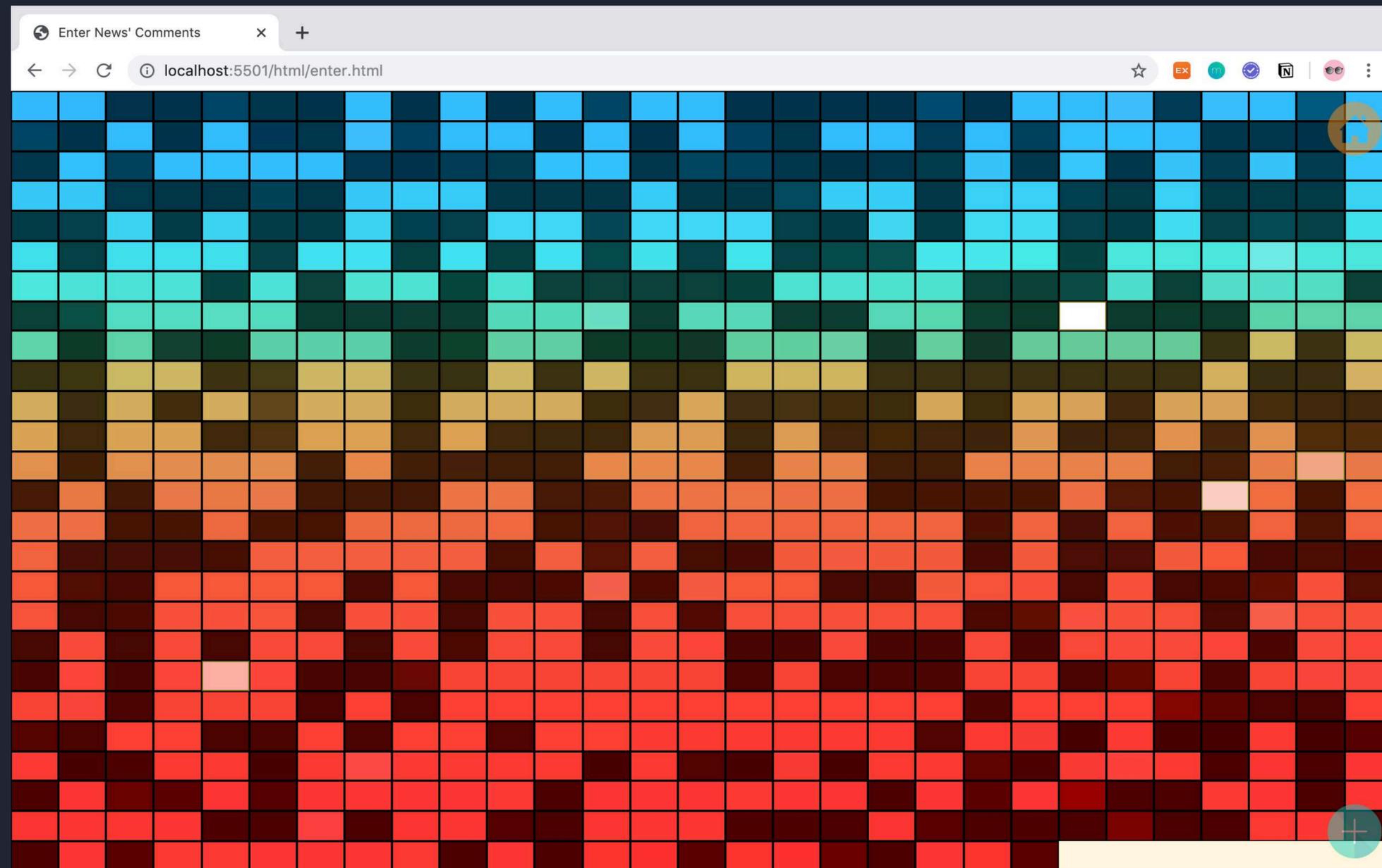


장점1.

한 기사에 대한 독자들의
전체적인 반응과 생각의 경향을
파악할 수 있다.

특징 3. 한 화면 내 전체 댓글 접근 가능

4가지 특징

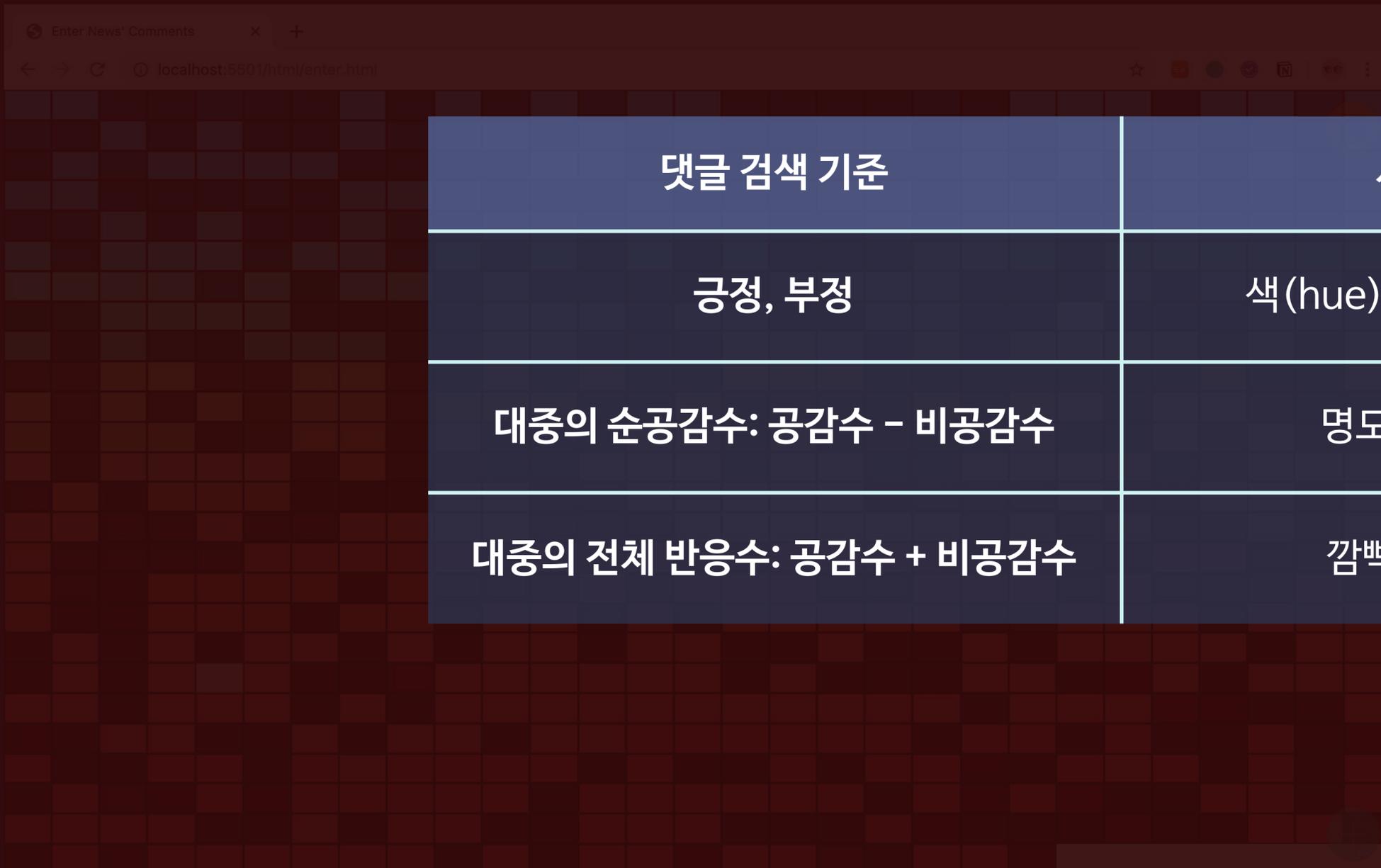


장점2.

직관적인 검색이 가능하다.

Solution

특징 3. 한 화면 내 전체 댓글 접근 가능

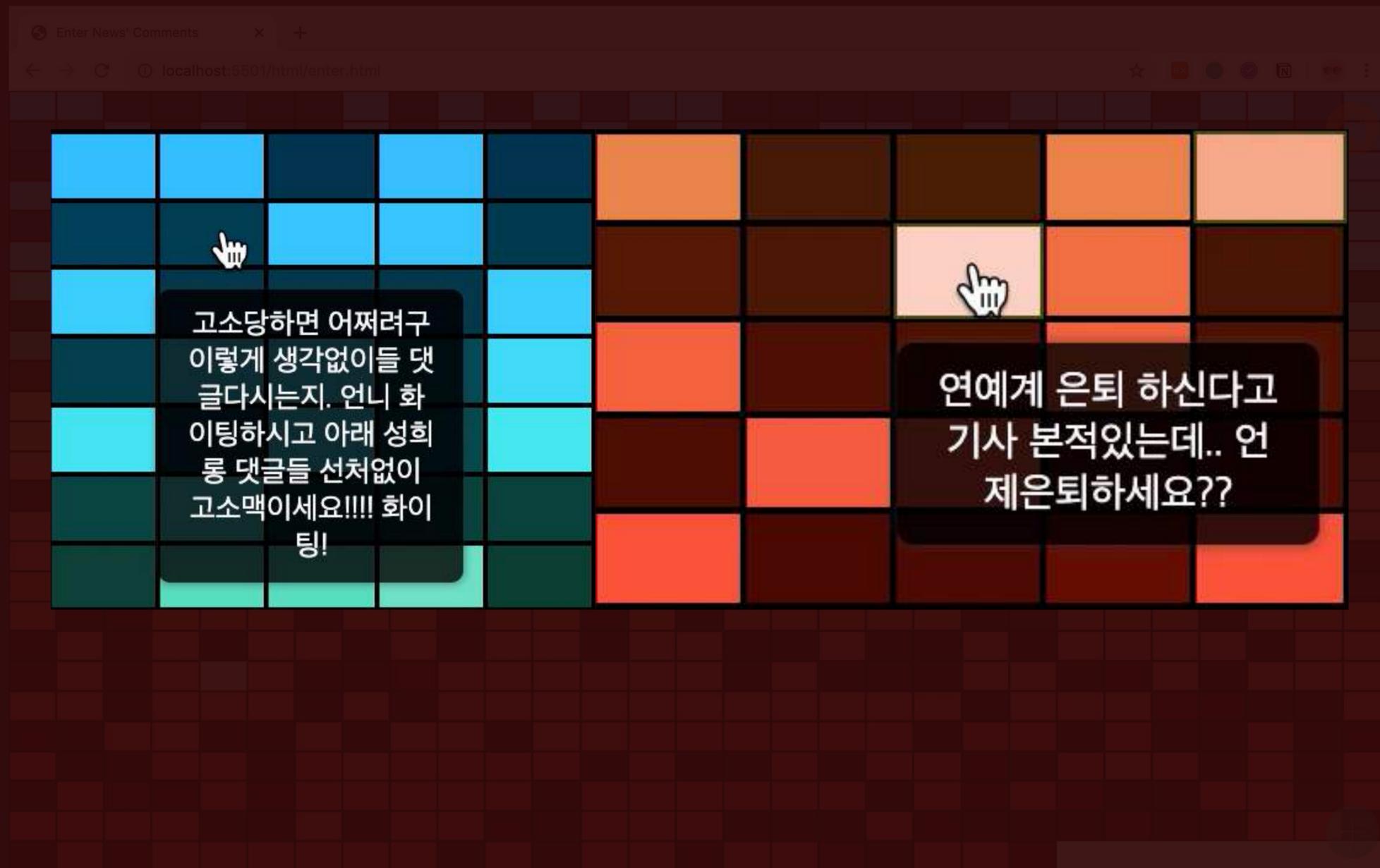


댓글 검색 기준	시각화 방법
긍정, 부정	색(hue), 채도(saturation)
대중의 순공감수: 공감수 - 비공감수	명도(brightness)
대중의 전체 반응수: 공감수 + 비공감수	깜빡임(blinking)

가능하다.

Solution

특징 3. 한 화면 내 전체 댓글 접근 가능

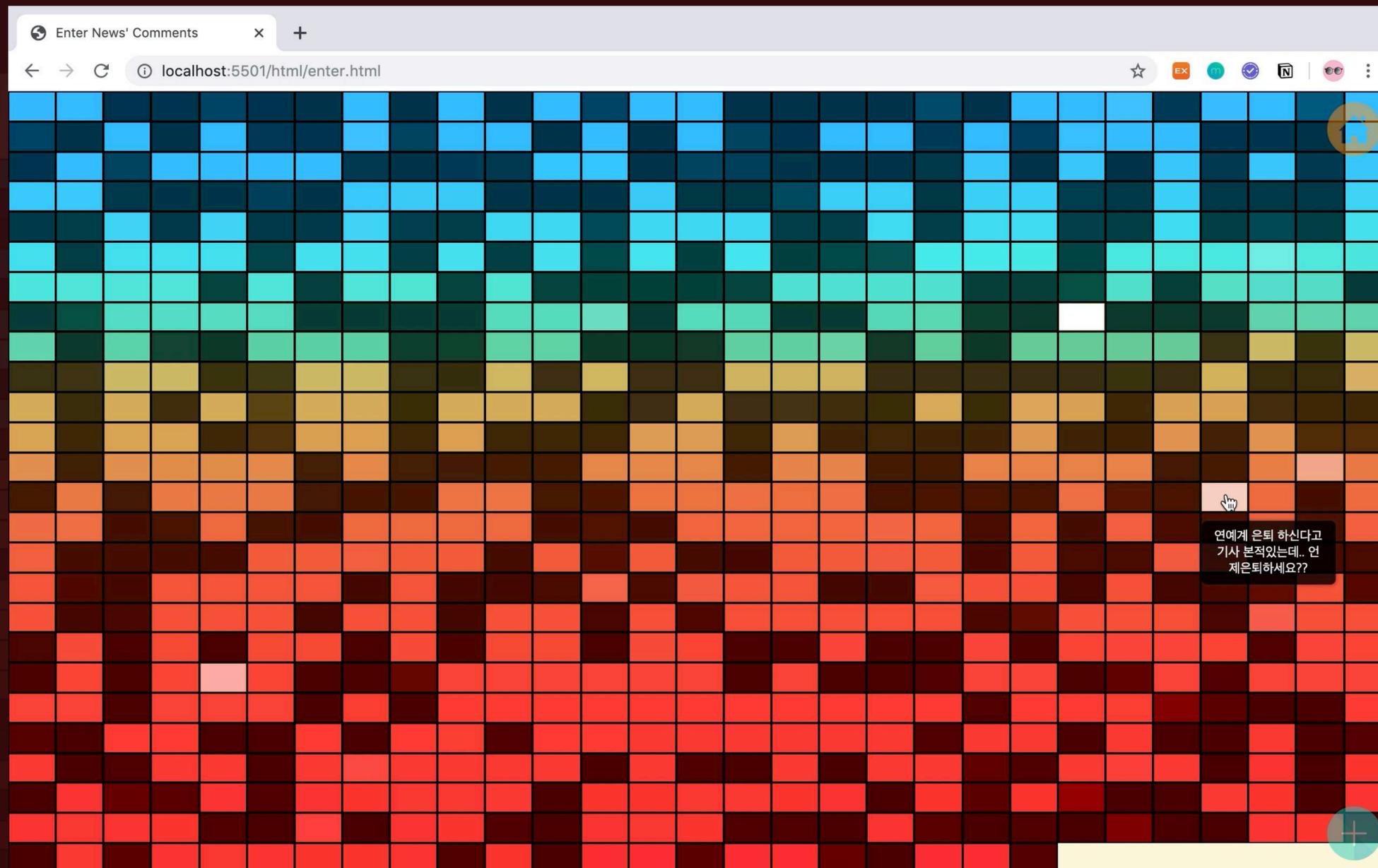


| 색조(hue) & 명도(saturation)

자정? 댓글의 색조와 명도를 통해 댓글이 긍정적인 댓글인지 부정적인 댓글인지와 공감을 많이 받고 있는지 비공감을 더 많이 받고 있는지를 알 수 있다.

Solution

특징 3. 한 화면 내 전체 댓글 접근 가능



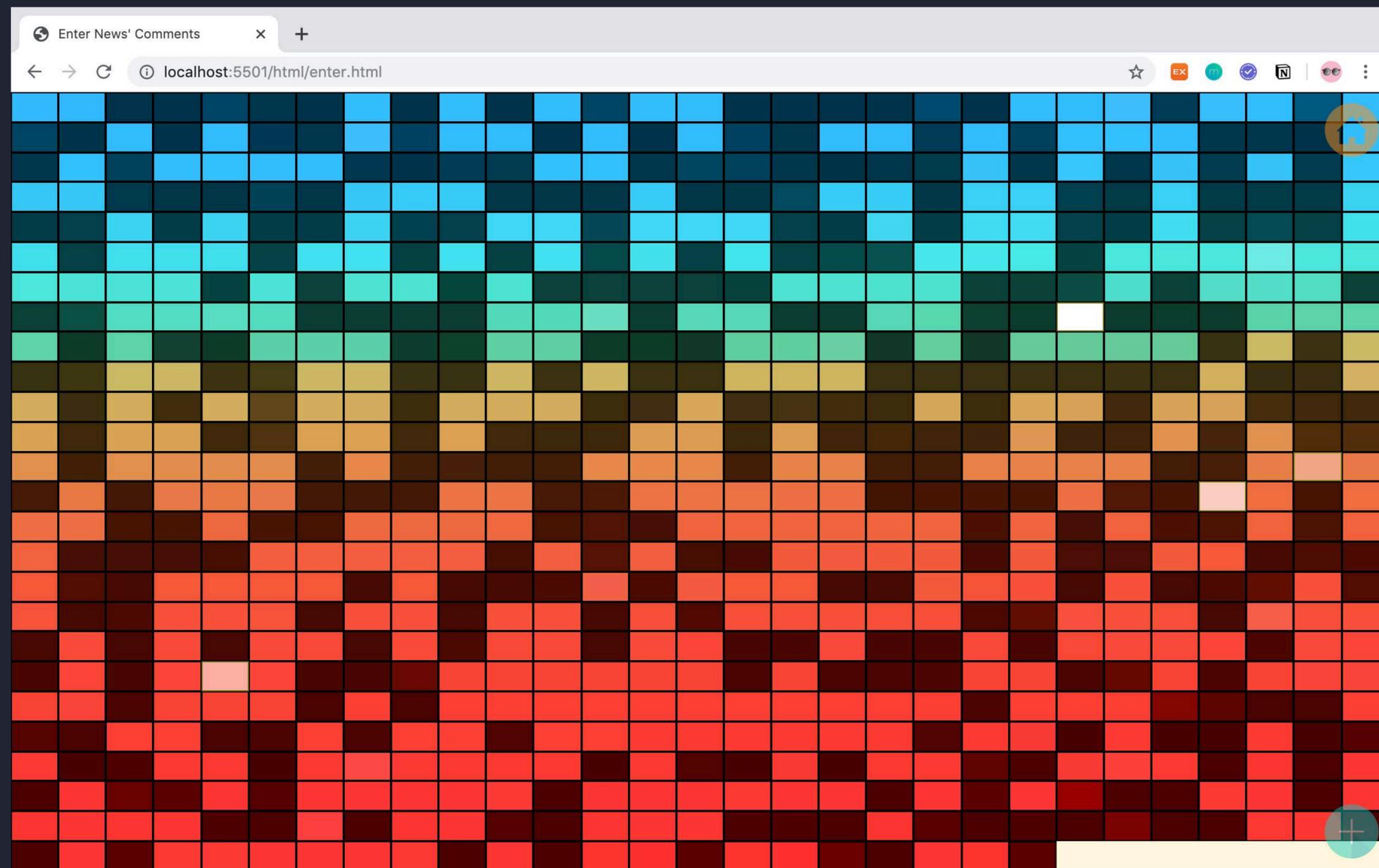
| 깜박임 (blinking)

특정 댓글이 사람들에게 반응을 많이 받으면 해당 사각형은 깜박임 (blinking) 효과를 갖게 되어 독자들은 어떤 댓글이 가장 관심을 많이 받고 있는지 직관적으로 검색이 가능하다.

Solution

특징 3. 한 화면 내 전체 댓글 접근 가능

4가지 특징



장점3.

독자들의 의견 개진 참여를 독려한다.

Solution

특징 4. 댓글 중심의 뉴스 기사 시각화

4가지 특징

댓글 검색 기준	시각화 방법
긍정, 부정	색(hue), 채도(saturation)
대중의 순공감수: $\text{공감수} - \text{비공감수}$	명도(brightness)
대중의 전체 반응수: $\text{공감수} + \text{비공감수}$	깜빡임(blinking)

뉴스 기사 시각화에 적용

Solution

특징 4. 댓글 중심의 뉴스 기사 시각화

4가지 특징

Feature.1

ANOP (All News on One Page)

Feature.4

댓글의 긍정/부정 시각화
기존 온라인 뉴스 사이트에서 사용하는 기사 열거 방식에서 탈피하여
특정 기간 동안의 모든 뉴스를 한 페이지에서 보여주는 방식

댓글 중심의
뉴스 기사 시각화

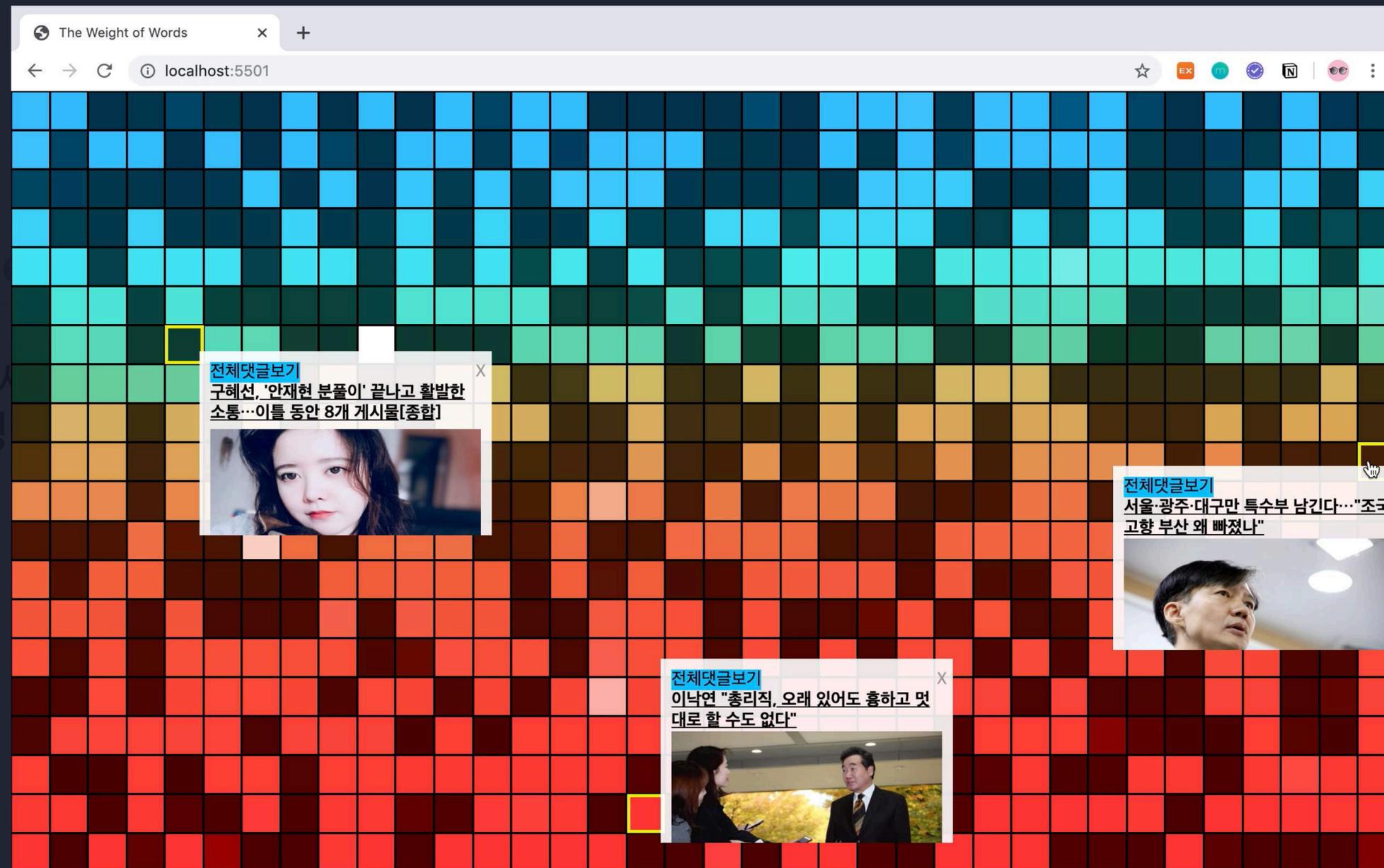
Solution

특징 4. 댓글 중심의 뉴스 기사 시각화

4가지 특징

| ANOP의 장점

1. 독자들이 최근 발생한 뉴스 기사들에 대해 어떤 반응을 보이고 있는지에 대한 경향을 직관적으로 파악할 수 있다.
2. 기사들이 긍정에서 부정으로 댓글 시각화와 같은 방식을 기준으로 기사를 정하여 읽을 수 있다.
3. 독자들에게 보다 많은 뉴스에 대한 접근성을 높여 생각의 폭을 넓힐 수 있는 기회를 제공한다.



Solution

INDEX

1. Introduction

서론

연구 배경 및 필요성

연구 목적

2. References

관련 연구 및 사례

악성 댓글 문제점 개선 사례

시각화 연구 사례

3. Solution

댓글 시각화 시스템

SG-Comment

뉴스 댓글 시스템

4. Evaluation

실험 및 평가

실험 방법

실험 결과

5. Discussion

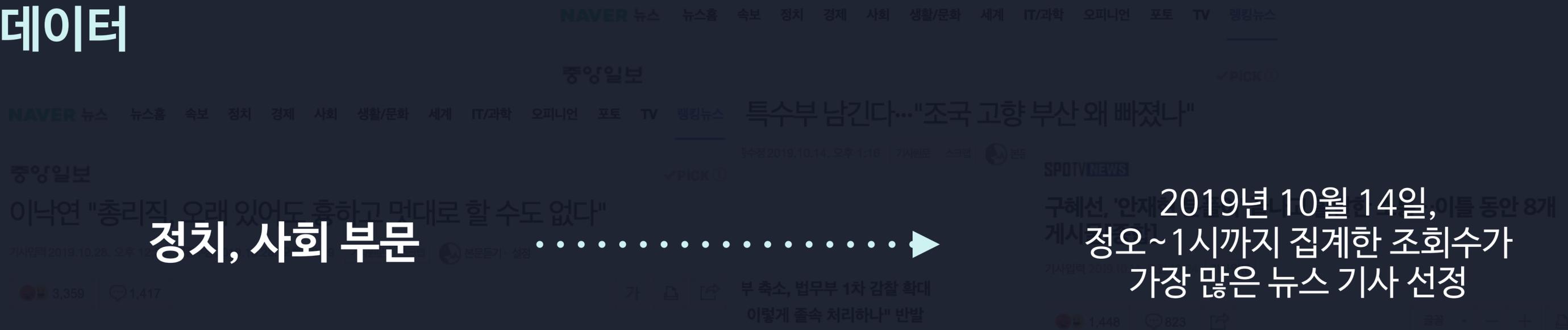
논의

6. Conclusion

결론

실험 방법

실험 데이터



정치, 사회 부문

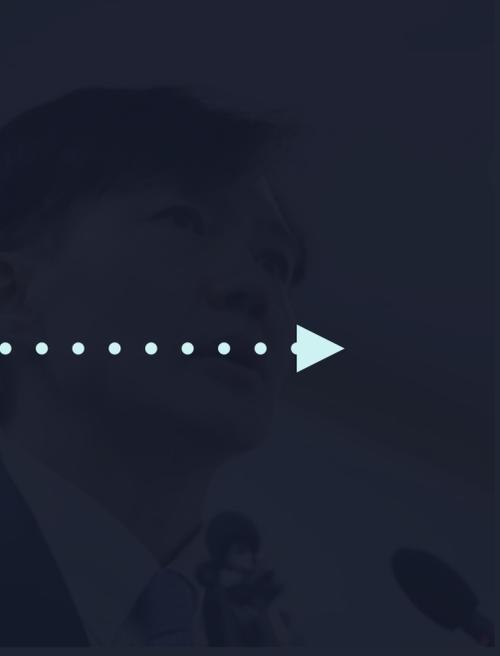


2019년 10월 14일, 정오~1시까지 집계한 조회수가 가장 많은 뉴스 기사 선정

"당 언제 돌아오나" 수도권 輿의원들 관심 이낙연 총리 국회 출석...본인 거취 물음에 답변



1987년 대통령 직선제 도입 이후 '최장수 총리' 기록을 세운 이낙연 국무총리가 28일 오전 정부서울청사로 출근하며 취재진의 질문에 답하고 있다. [연합뉴스]



2천 경기도 정부과천청사에서 특수부 축소 등을 주요 내용으로

2019년 10월 14일, TV 연예뉴스에서 일간 많이 본 뉴스 중 랭킹 1위 뉴스 기사 선정

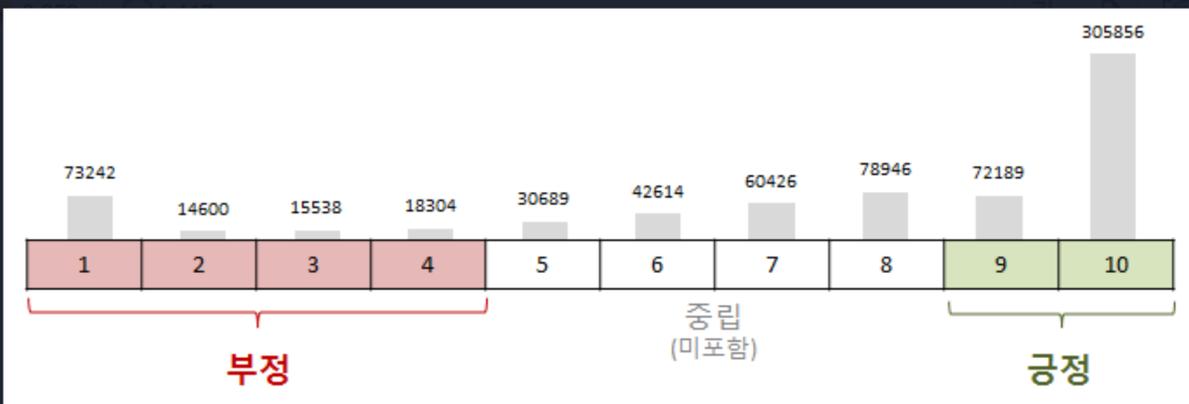
구혜선, '안재현 분풀이' 끝나고 활발한 소통 ... 이틀 동안 8개 게시물 [종합]

구혜선이 안재현 분풀이 ... 이틀 동안 8개 게시물 [종합]

Evaluation

실험 방법

공정/부정 머신 러닝 모델



부 축소, 법무부 1차 검찰 확대
이렇게 줄속 처리

영화당 100개의 리뷰를 모아 총 20만개 리뷰로 이루어져 있고, 1~10까지의 평점 중에서 중립적인 평점(5~8)은 제외하고 1~4점을 긍정으로, 9~10점을 부정으로 동일한 비율로 샘플링 되어 있는 네이버 영화 평점 및 리뷰 데이터셋 사용

Cyclam3n's Blog. (n.d.). [online] Available at: https://cyclam3n.github.io/2018/11/10/classifying_korean_movie_review.html [Accessed 11 Aug. 2019]

Evaluation

1987년 대통령 직선제 도입 이후 '리간수 총리' 기록을 세운 이낙연 국무총리가 28일 오전 정부서울청사로 출근하며 취재진의 질문에 답하고 있다. (연합뉴스)

1천 경기도 정부과천청사에서 특수부 축소 등을 주요 내용으로

구혜선, '안재현 분풀이' 끝나고 활발한 소통...이틀 동안 8개 게시물[종합]

실험 방법

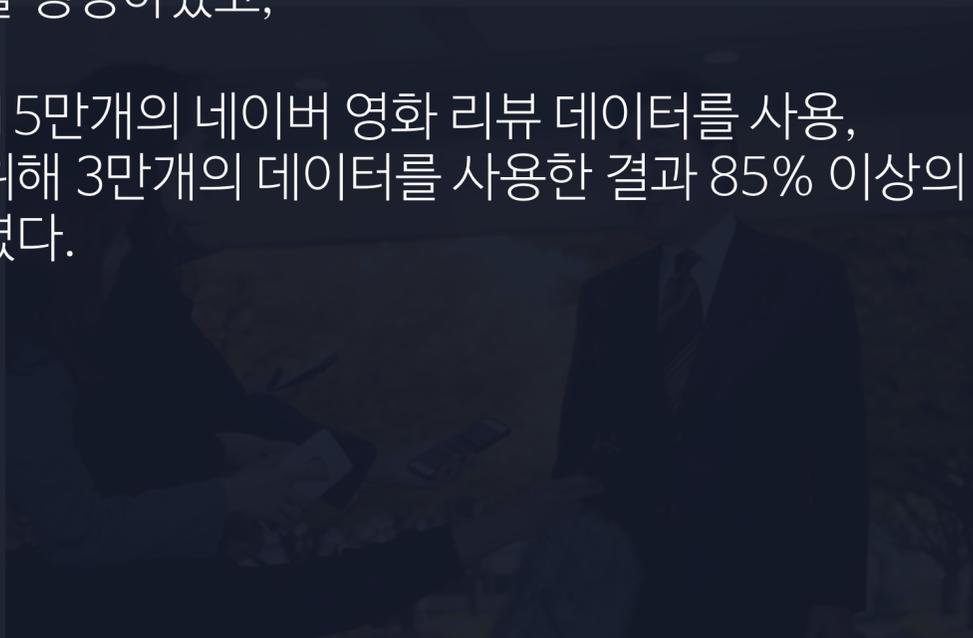
긍정/부정 머신 러닝 모델

| 케라스(Keras) 모델을 적용한 데이터 셋

파이썬으로 구현된 딥러닝 라이브러리인 케라스(Keras)를 활용하여 댓글의 긍정/부정, 퍼센티지를 예측할 수 있는 모델을 만들었다.

가장 빈도 수가 높은 1만개의 단어에 대해 Count Vectorization 기법으로 각 댓글 문장을 1만 차원의 벡터로 매핑하여 학습 데이터를 생성하였고,

학습을 위해 15만개의 네이버 영화 리뷰 데이터를 사용, 성능평가를 위해 3만개의 데이터를 사용한 결과 85% 이상의 정확도를 보였다.



1987년 개봉된 리전제 도입 이후 '리전제'를 기록한 세운 이낙연 국무총리가 28일 오전 정부서울청사로 출근하며 리전제의 침묵에 담고 있다. (연합뉴스)

NAVER 뉴스 뉴스홈 속보 정치 경제 사회 생활/문화 세계 IT/과학 오피니언 포토 TV 팽킹뉴스

중앙일보

NAVER 뉴스 뉴스홈 속보 정치 경제 사회 생활/문화 세계 IT/과학 오피니언 포토 TV 팽킹뉴스

comments	like	dislike	predict	percent	truelike	wholelike
토끼	3416	381	1	60.58	3035	3797
안녕하세요	1782	146	0	85.11	1636	1928
안녕하세요	1483	137	0	95.9	1346	1620
안녕하세요	922	96	0	80.43	826	1018
안녕하세요	527	28	1	77.18	499	555
안녕하세요	350	29	1	65.76	321	379
안녕하세요	241	24	0	93.5	217	265
안녕하세요	197	7	0	98.49	190	204
안녕하세요	192	22	0	91.65	170	214
안녕하세요	156	14	1	54.52	142	170
안녕하세요	129	7	1	99.89	122	136
안녕하세요	185	64	1	99.47	121	249
안녕하세요	127	10	0	99.76	117	137
안녕하세요	108	1	0	98.95	107	109
안녕하세요	122	26	0	59.4	96	148
안녕하세요	89	17	0	91.71	72	106
안녕하세요	81	9	1	99.86	72	90
안녕하세요	76	9	1	68.31	67	85
안녕하세요	60	3	0	99.53	57	63
안녕하세요	62	7	0	93.37	55	69
안녕하세요	52	5	1	76.68	47	57

Evaluation

실험 방법

긍정/부정 머신 러닝 모델

| 케라스(Keras) 모델을 적용한 데이터 셋

파이썬으로 구현된 딥러닝 라이브러리인 케라스(Keras)를 활용하여 댓글의 긍정/부정, 퍼센티지를 예측할 수 있는 모델을 만들었다.

가장 빈도 수가 높은 1만개의 단어에 대해 Count Vectorization 기법으로 각 댓글 문장을 1만 차원의 벡터로 매핑하여 학습 데이터를 생성하였고,

학습을 위해 15만개의 네이버 영화 리뷰 데이터를 사용, 성능평가를 위해 3만개의 데이터를 사용한 결과 85% 이상의 정확도를 보였다.

‘힘내세요~ 세상은 넓고 아름답습니다’ 라는 댓글은 99.79%로 긍정 댓글로 분류
‘재수없으니~ 나오지마라..튀튀튀’ 라는 부정 댓글은 96.66%의 확률로 부정댓글로 분류

comments	like	dislike	predict	percent	truelike	wholelike
힘내세요~ 세상은 넓고 아름답습니다	3416	381	1	60.58	3035	3797
재수없으니~ 나오지마라..튀튀튀	1782	146	0	85.11	1636	1928
힘내세요~ 세상은 넓고 아름답습니다	1483	137	0	95.9	1346	1620
재수없으니~ 나오지마라..튀튀튀	922	96	0	80.43	826	1018
힘내세요~ 세상은 넓고 아름답습니다	527	28	1	77.18	499	555
재수없으니~ 나오지마라..튀튀튀	350	29	1	65.76	321	379
힘내세요~ 세상은 넓고 아름답습니다	241	24	0	93.5	217	265
재수없으니~ 나오지마라..튀튀튀	197	7	0	98.49	190	204
힘내세요~ 세상은 넓고 아름답습니다	156	14	1	54.52	142	170
재수없으니~ 나오지마라..튀튀튀	129	7	1	99.89	122	136
힘내세요~ 세상은 넓고 아름답습니다	127	10	0	99.76	117	137
재수없으니~ 나오지마라..튀튀튀	108	1	0	98.95	107	109
힘내세요~ 세상은 넓고 아름답습니다	122	26	0	59.4	96	148
재수없으니~ 나오지마라..튀튀튀	89	17	0	91.71	72	106
힘내세요~ 세상은 넓고 아름답습니다	81	9	1	99.86	72	90
재수없으니~ 나오지마라..튀튀튀	76	9	1	68.31	67	85
힘내세요~ 세상은 넓고 아름답습니다	60	3	0	99.53	57	63
재수없으니~ 나오지마라..튀튀튀	62	7	0	93.37	55	69
힘내세요~ 세상은 넓고 아름답습니다	52	5	1	76.68	47	57

Evaluation

실험 방법

실험 대상 선정 및 과정

온라인을 통해 모집한 10~30대, 총 100명을 대상으로 설문조사를 진행하였다.
 실험 집단은 네이버 뉴스 기사를 일주일에 한 건 이상 보는 사람들을 대상으로 선정하였다.

나이	A 그룹	B 그룹
15 - 19세	15	12
20 - 29세	24	27
30 - 39세	11	11
그룹 별 참가 인원	50	50
총 참가 인원	100 명 (남 12명, 여 88명)	

Evaluation

실험 방법

실험 대상 선정 및 과정

네이버 뉴스 댓글 시스템과 *SG-Comment* 시스템 간 비교를 위한 설문 문항

No.	설문 항목
A-1	댓글 시스템을 통해 기사 내용에 대해 많은 독자들의 생각을 쉽게 확인할 수 있었다.
A-2	댓글 시스템을 통해 기사 내용에 대해 전반적인 독자들의 긍정/부정 반응 경향을 쉽게 파악할 수 있었다.
A-3	댓글을 작성할 때, 내가 남기는 댓글이 타인에게 미칠 수 있는 영향에 대해 생각하였다.
A-4	댓글 시스템에서 악성 댓글이 쉽게 노출될 것 같다.
A-5	뉴스 홈 시스템을 통해 많은 뉴스 기사에 쉽게 접근할 수 있었다.

SG-Comment 시스템에만 관련된 추가 설문 항목

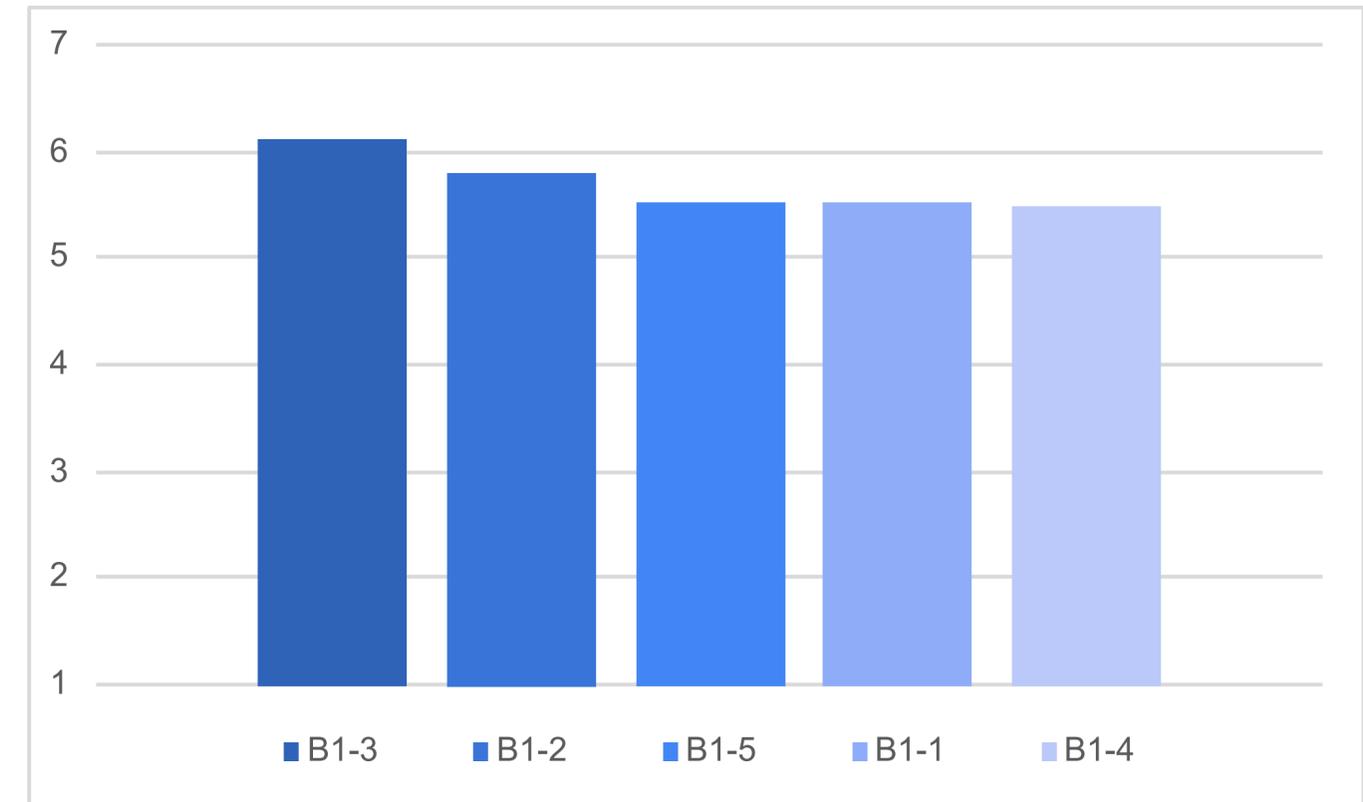
No.	설문 항목
B-1	한 페이지에서 세가지 기준으로 - 공부정순(색깔), 순공감순(밝기), 전체공감순(깜빡임) - 댓글을 검색하는 것이 도움이 되었다.
B-2	뉴스 홈에서 특정 시간에 사람들이 어떤 기사에 관심을 가지는지, 각 기사에 달린 전체 댓글 수와 연동된 깜빡이는 시각화와 아울러 그 댓글들의 평균 긍정/부정 정도를 색깔로 시각화한 것이 기사를 검색하는 데 도움이 되었다.
B-3	뉴스 홈 시스템을 통해 전체 기사에 대한 독자들의 긍정/부정 반응 경향을 쉽게 파악할 수 있었다.
B-4	이 댓글 시스템은 댓글 문화를 개선하는데 도움이 될 것 같다.
B-5	이 뉴스 시스템을 다시 사용하고 싶다.

실험 결과

정량 평가 결과

5가지 사용자 경험 질문 항목에 대한 네이버 뉴스 댓글 시스템 vs SG-Comment 대응표본 t-검정 결과

Questions	Naver News Comments	SG-Comment	t	p	dF
A-1	4.92 ± 1.28	5.82 ± 1.01	-6.25	< 0.001***	99
A-2	4.88 ± 1.62	5.82 ± 1.30	-5.10	< 0.001***	99
A-3	5.32 ± 1.57	5.92 ± 1.09	-4.24	< 0.001***	99
A-4	5.89 ± 1.12	5.49 ± 1.54	2.28	.025*	99
A-5	5.52 ± 1.11	5.66 ± 1.15	-1.04	.302	99



Evaluation

본 논문에서 제안하는 방식으로 댓글을 시각화 하는 것에 대해서 통계적으로 유의한 차이를 보였으나, 뉴스 홈 화면을 시각화 하는 것에 대해서는 통계적으로 유의한 차이가 나타나지 않았다.

실험 결과

정성 평가 결과

“댓글을 보는 이유가 전체적으로 사람들이 해당 기사에 대해 어떤 생각을 가지는지, 어떠한 감정을 나타내고 있는지 살펴보기 위해 댓글을 읽는데 색과 채도, 그리고 명도로 **댓글을 직관적으로 볼 수 있게 하는 부분에서** 도움이 되었다. 또한, 서로 다른 의견이 충돌할 때, 상대방의 입장도 볼 수 있어서 **완충재 역할을 하여 댓글 문화를 개선하는데** 도움이 될 것 같다”

(23세, 여성)

“색으로 긍정적인 댓글과 부정적인 댓글로 나뉘져 있어 **보고 싶은 성향의 댓글만 볼 수 있어서** 보기 편했다. 악성 댓글을 볼 때 느꼈던 **피로도를 줄이는데** 도움이 되었다.”

(20세, 여성)

실험 결과

정성 평가 결과

“부정적인 댓글을 작성할 때 사이렌 아이콘이 나타나서 중간에 작성한 댓글을 다시 보게 되었다. 그 이후로 댓글을 쓸 때마다 신경을 써서 댓글을 작성하게 되었다.”

(28세, 남성)

“댓글은 색깔로 시각화해서 보여주는게 도움이 된다고 생각하는데 뉴스 홈 화면에서는 내용 위주로 노출이 되면 좋을 것 같다. 사람들의 반응 보다는 기사에 대한 정보를 바탕으로 시각화하면 좋을 것 같다.”

(26세, 여성)

INDEX

1. Introduction

서론

연구 배경 및 필요성

연구 목적

2. References

관련 연구 및 사례

악성 댓글 문제점 개선 사례

시각화 연구 사례

3. Solution

댓글 시각화 시스템

SG-Comment

뉴스 댓글 시스템

4. Evaluation

실험 및 평가

실험 방법

실험 결과

5. Discussion

논의

6. Conclusion

결론

논의 결과

정량 평가 결과

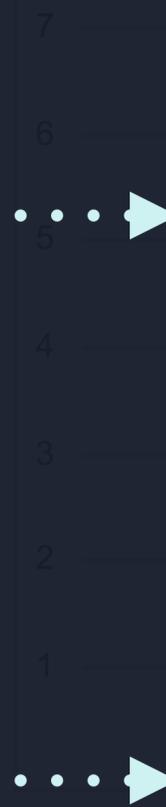
5가지 사용자 경험 질문 항목에 대한 네이버 뉴스 댓글 시스템 vs SG-Comment 대응표본 t-검정 결과

Questions	Naver News Comments	SG-Comment	t	p	dF
A-1	4.92 ± 1.28	5.82 ± 1.01	-6.25	< 0.001***	99
A-2	4.88 ± 1.62	5.82 ± 1.30	-5.10	< 0.001***	99
A-3	5.32 ± 1.57	5.92 ± 1.09	-4.24	< 0.001***	99
A-4	5.89 ± 1.12	5.49 ± 1.54	2.28	.025*	99
A-5	5.52 ± 1.11	5.66 ± 1.15	-1.04	.302	99

SG-Comment

시스템

ANOP



인터랙티브한 피드백이 간단하면서도 큰 실효성이 있을 수 있음을 보여주지만, 이러한 피드백이 반복될 경우 효과가 경감될 수 있어 장기적인 사용에 따른 효과성 검증이 필요하다.

댓글과 기사는 서로 다른 특성을 지니고, 사용자가 원하는 것이 다르기 때문에 댓글에 적용된 시각화를 확장하는 것은 효과적이지 않다.

본 논문에서 제안하는 방식으로 댓글을 시각화 하는 것에 대해서 통계적으로 유의한 차이를 보였으나, 뉴스 홈 화면을 시각화 하는 것에 대해서는 통계적으로 유의한 차이가 나타나지 않았다.

Discussion

논의

연구 제한점

- 제한점 -

1. 댓글의 긍정, 부정을 판단하는 **모델의 정확도 개선 필요**
2. 뉴스 플랫폼을 사용하는 50~60대 등의 **다른 높은 연령대로 결과를 일반화 시키기는 어려움**
3. 데스크 탑 컴퓨터 사용에 최적화되어 있어 **모바일 기기에서의 사용은 별도 연구 필요**

INDEX

1. Introduction

서론

연구 배경 및 필요성

연구 목적

2. References

관련 연구 및 사례

악성 댓글 문제점 개선 사례

시각화 연구 사례

3. Solution

댓글 시각화 시스템

SG-Comment

뉴스 댓글 시스템

4. Evaluation

실험 및 평가

실험 방법

실험 결과

5. Discussion

논의

6. Conclusion

결론

결론

SG-Comment (Self-reGulating Comment)

머신 러닝 기반으로 댓글 작성 중 댓글의 긍정부정도를 실시간으로 판단하여 그에 따른 시각화를 제공함으로써 자기 검열의 기회를 제공하고, 색상과 시간 지연 방식으로 악성 댓글의 노출도를 줄이며, 한 페이지에서 모든 댓글을 직관적으로 검색할 수 있는 기능을 특징으로 한다.



온라인 문화에서 악성 댓글을 줄이고 건전한 댓글 문화를 만드는데 대안적인 시스템으로 사용

박지현 | Jihyun Park

carpedie7720@gmail.com

<https://www.linkedin.com/in/jihyuns/>

<http://jihyuns.com/>

기술적 조치를 통한 혐오 표현 대응
악성댓글 처리 알고리즘을 활용한 댓글 시각화