


과제 유형	일반 캡스톤디자인					
캡스톤디자인(종합설계) 과제 결과 보고서						
개설학과	컴퓨터공학과			학기	<input type="checkbox"/> 1학기 <input type="checkbox"/> 1·2학기 연계 <input checked="" type="checkbox"/> 2학기 <input type="checkbox"/> 하계 <input type="checkbox"/> 동계	
교과목명	캡스톤디자인II			학점	3	
과제명	자연어 처리 기반 영화 리뷰 필터링 및 시각화					
지도교수	소속(학과)	컴퓨터공학과			성명	황경호
구분	소속(학과)	학번	학년	성명	연락처	E-mail
참여학생 (총명)	컴퓨터공학과	20171615	4	서민석	010-2856-4834	seominseok4834@gmail.com
	컴퓨터공학과	20171602	4	한재혁	010-5584-9805	hjh0633@gmail.com
	컴퓨터공학과	20171598	4	정인상	010-9620-1518	jisk101614@gmail.com
과제요약	국내 영화 사이트에서 리뷰 데이터를 크롤링해 사용자에게 도움이 되지 않는 리뷰들을 필터링한다. 필터링을 통해 검증된 리뷰들은 감상 포인트 추출 모델을 사용해 텍스트 내에 있는 관람 포인트에 대한 평가를 분석한다. 필터링된 리뷰와 감상 포인트 분석 결과는 어플리케이션을 통해 보여준다.					
위와 같이 캡스톤디자인(종합설계) 과제를 수행하였으며, 결과보고서를 제출합니다. <div style="text-align: center;">2022. 12 . 5 .</div> <div style="display: flex; justify-content: space-between; margin-top: 20px;"> <div>지도교수 :</div> <div>황경호 (인)</div> </div> <div style="display: flex; justify-content: space-between; margin-top: 20px;"> <div>대표학생 :</div> <div>한재혁 </div> </div>						
한밭대학교 산학협력교육원장 귀하						

캡스톤디자인(종합설계) 과제 수행 보고서

1. 과제(주제) 선정 배경 및 필요성

리뷰는 상품에 대한 일반 정보를 포함하여 실사용자들의 주관적인 후기가 들어있어 소비자들이 상품을 소비하는데 있어 중요한 지표가 된다. 하지만 이러한 점을 악용하는 바이럴 마케팅으로 인해 리뷰에 대한 소비자들의 불신도가 깊어졌다.

악성 및 허위 리뷰에 대한 국내외 연구나 산업을 보면 아마존과 배달의 민족, 왓차피디아 등 많은 기업들이 인공지능 기술 및 자체적으로 개발한 알고리즘을 활용하여 실시간 필터링 및 모니터링으로 리뷰들을 차단하고 있다.

그러나 기업들의 이러한 노력에도 리뷰어들과 주관적인 내용이 담긴 악성 리뷰를 모두 걸러내기에는 한계가 있다. 일례로 국내 최대 사용자를 보유한 네이버 영화의 경우 실관람객이 아니더라도 리뷰를 작성할 수 있어 댓글 알바나 허위 리뷰에 의해 정상적인 소비활동에 차질이 생길 수 있다. 때문에 이를 보완하기 위한 기술이 아직까지도 연구, 개발되고 있으며 해당 기술의 완성도가 높아진다면 소비자들의 질 높은 소비 활동이 가능해질 것으로 보인다.

본 과제에서는 여러 영화 리뷰 사이트의 리뷰에서 신뢰할만한 리뷰를 추출하는 것을 최우선 목표로 세웠다. 또한 해당 리뷰에서 관람객들이 만족했거나 아쉬웠던 감상 포인트를 그래프로 시각화해 검증된 리뷰와 함께 제공하는 것이 프로젝트의 목표이다.

2. 과제(주제) 설계 내용 및 추진 방법

국내 주요 영화 리뷰 사이트에서 리뷰 정보를 크롤링해 편향적인 리뷰, 근거없는 비판, 맹목적인 칭찬, 비관람객이 작성한 리뷰, 내용이 없는 리뷰를 필터링해서 신뢰성 있는 리뷰를 획득한다. 크롤링한 리뷰에서 속성 기반 감성 분석을 진행해 해당 영화의 감상 포인트별 긍정 비율을 계산해 웹/앱 애플리케이션을 통해 제공할 예정이다.

또한 오픈 API를 통한 영화 검색 기능과 영화 정보 제공 기능, 자연어 처리를 기반으로 진행되는 리뷰 필터링과 감상 포인트 시각화이며 이는 SPA 방식으로 구현된 웹/앱 애플리케이션을 통해 보여질 예정이다.

3. 과제(주제) 활동 결과 및 결과물(시제품)

1. 웹/앱 애플리케이션

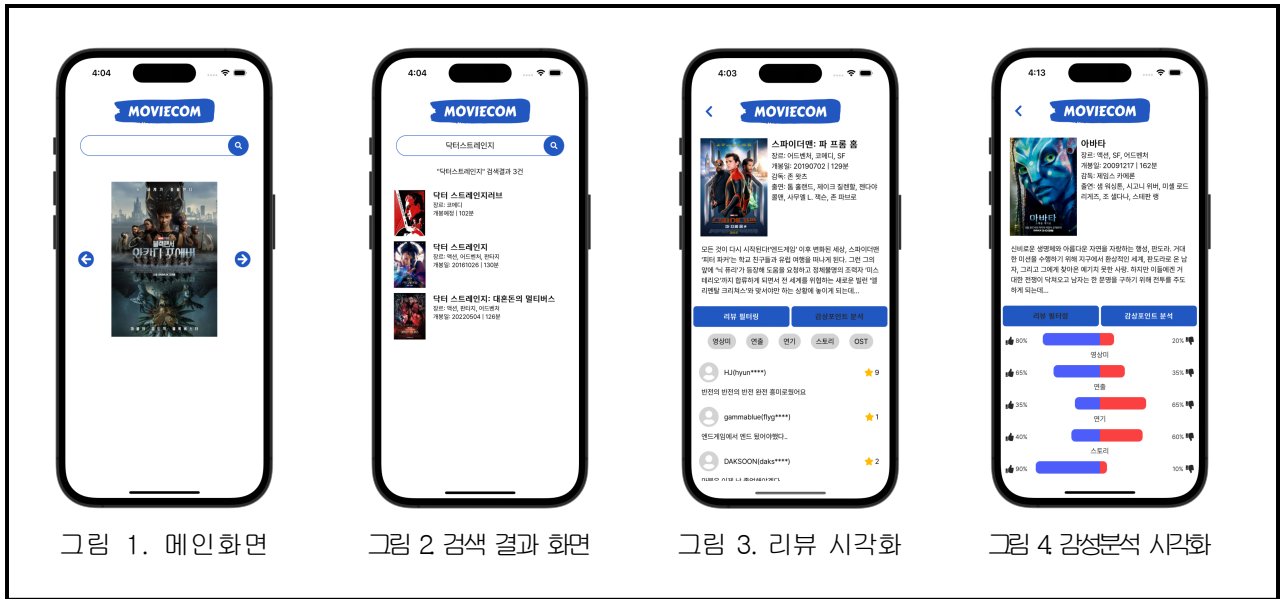
- React를 이용한 웹 애플리케이션이 제작 완료
- React Native를 이용해 웹 애플리케이션을 앱 형태로 전환 완료
- Custom API를 위한 API 서버 및 DB 구축 완료
- 네이버 영화의 실관람객의 리뷰 데이터 크롤링 구현 완료

2. 인공지능 모델

- 속성 기반 감성 분석을 이용한 리뷰 필터링 및 감상 포인트 추출 모델 개발 완료

3. 데이터 구축

- 리뷰 필터링을 위한 라벨링 데이터 740개 구축
- 감상 포인트 추출을 위한 라벨링 데이터 1,326개 구축



5. 결과물의 활용방안(사업화 연계방안)

다양한 상품에 대한 리뷰가 넘쳐나는 현재, 리뷰 데이터를 분석해 소비자의 만족도를 끌어올리는 것은 많은 이커머스 기업들의 목표이다. 속성 기반 감성분석은 기존 감성 분석 모델보다 발전된 형태의 감성 분석 기술로 국내에서도 관련 연구가 이루어지고 있지만 현재까지는 적극적으로 활용되고 있지 않다. 본 캡스톤 디자인은 영화 리뷰에만 이를 적용했지만 이에 국한되지 않고, 다양한 소비 영역으로 확장할 수 있을 것이라 생각된다.

캡스톤 디자인 II 최종결과 보고서

프로젝트 제목(국문): 자연어 처리 기반 영화 리뷰 필터링 및 시각화

프로젝트 제목(영문): Filter and visualize movie reviews based on NLP

프로젝트 팀(원): 학번: 20171601 이름: 한재혁

프로젝트 팀(원): 학번: 20171615 이름: 서민석

프로젝트 팀(원): 학번: 20171598 이름: 정인상

1. 중간보고서의 검토결과 심사위원의 '수정 및 개선 의견'과 그러한 검토의견을 반영하여 개선한 부분을 명시하시오.

- 영화를 선택하면 실시간으로 해당 영화에 대한 리뷰 데이터를 크롤링 한 뒤 필터링 모델 및 감상 포인트 추출 모델을 사용하여 분석한 결과를 DB에 저장하여 이후 요청시 바로 반환한다.
- 데모 사이트를 추가로 제작하여 사용자가 리뷰 입력시 필터링 모델 및 감상 포인트 추출 모델 분석 결과를 시각화하여 보여준다.

2. 기능, 성능 및 품질 요구사항을 충족하기 위해 본 개발 프로젝트에서 적용한 주요 알고리즘, 설계방법 등을 기술하시오.

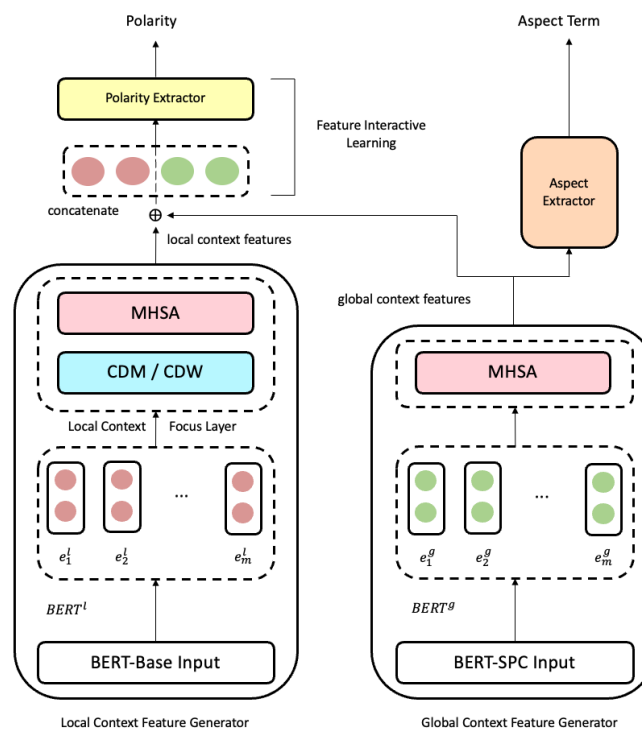


그림 6. LCF-ATEPC 모델

속성 기반 감성 분석(Aspect Based Sentiment Analysis)은 문장이나 문서 내에 있는 여러 aspect에 대한 독립적인 감성을 예측할 수 있다.

속성 기반 감성 분석은 텍스트 안에 포함된 속성어를 추출하는 aspect term extraction task와 속성어에 대한 감성을 예측하는 aspect polarity classification task로 구성돼 있는데, 본 캡스톤 디자인에서 사용한 local context focus atepc 모델은 Local Context Feature Generator(LCFG)에서 문맥 단어의 특징을 보존하며, 의미상으로 상대적인 단어의 피처를 마스킹하거나 가중치를 약화해 local context feature를 생성한다. 이후 Global Context Feature Generator(GCFG)에서 생성된 global context feature와 결합해 두 task를 학습한다.

3. 요구사항 정의서에 명시된 기능 및 품질 요구사항에 대하여 최종 완료된 결과를 기술하시오.



그림 2. 분석 결과를 확인해 볼 수 있는 데모 사이트

사용자가 리뷰를 입력하면 리뷰 필터링 및 감상 포인트 추출 모델을 통해 분석한 결과를 확인할 수 있는 데모 사이트로, 필터링 된 리뷰는 빨간색 배경의 바운딩 박스로 표시되며, 오른쪽에 나타나게 된다. 필터링 되지 않은 리뷰들은 파란색 배경의 바운딩 박스로 표시되고, 감상 포인트 추출 모델을 사용하여 해당 텍스트 내에 있는 감상 포인트들을 추출하고, 해당 평가가 긍정인지 부정인지를 분석해 긍정일 경우 초록색, 부정일 경우 분홍색으로 하이라이팅 한다.

4. 구현하지 못한 기능 요구사항이 있다면 그 이유와 해결방안을 기술하시오,

최초 요구사항	구현 여부(미구현, 수정, 삭제 등)	이유(일정부족, 프로젝트 관리미비, 팀원변동, 기술적 문제 등)
리뷰 필터링 모델 개발	구현	데이터 구축 문제로 인해 요구사항과 달리 1가지 기준에 대해서만 필터링하도록 구현
감상 포인트 추출 모델 개발	구현	해당 사항 없음

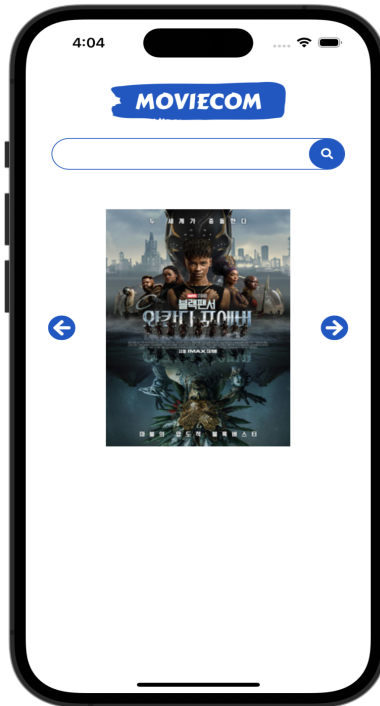
5. 요구사항을 충족시키지 못한 성능, 품질 요구사항이 있다면 그 이유와 해결방안을 기술하시오.

분류(성능, 속도 등) 및 최초 요구사항	충족 여부(현재 측정결과 제시)	이유(일정부족, 프로젝트 관리미비, 팀원변동, 기술적 문제 등)
리뷰 필터링 모델 성능	측정 불가	5가지 기준이 아닌 1가지 기준에 대해서만 필터링하기 때문에 측정 불가
감상 포인트 추출 모델 성능	APC-Task F1 Score: 93.51 ATE-Task F1 Score: 83.15	데이터셋 부족으로 인한 문제로, 추가적으로 데이터를 구축할 시 해결될 것으로 보임

6. 최종 완성된 프로젝트 결과물(소프트웨어, 하드웨어 등)을 설치하여 사용하기 위한 사용자 매뉴얼을 작성하시오.



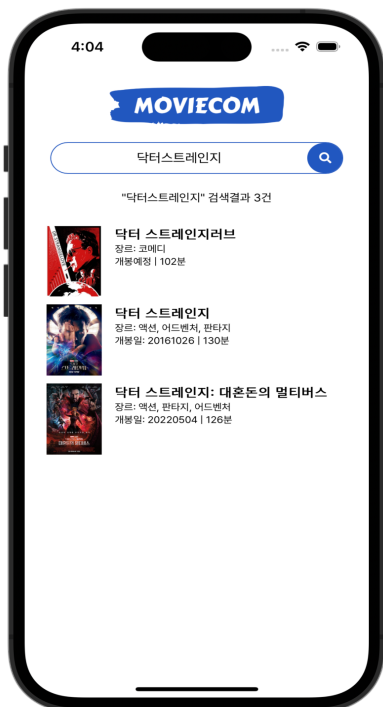
1. 어플리케이션 구동시 화면



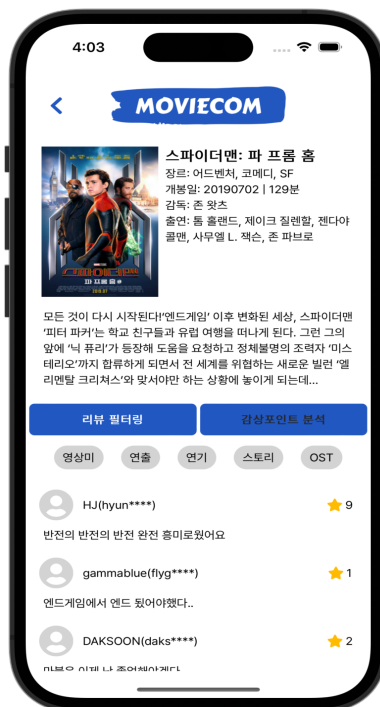
2. 초기 화면



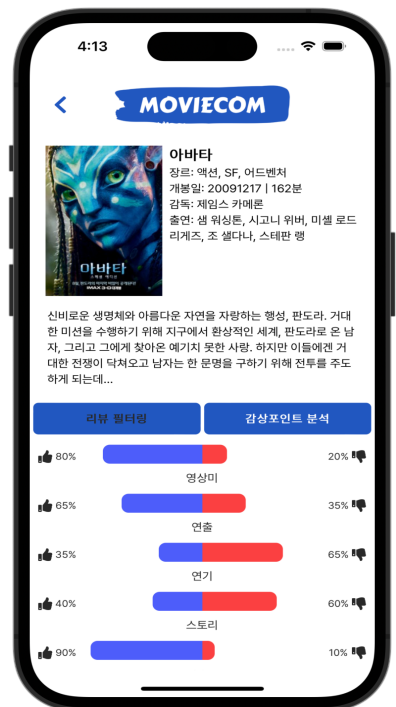
2-1. 박스오피스 영화 선택시 화면



3. 영화 검색 화면



4. 영화 리뷰 확인 화면



5. 감상 포인트 분석 결과 확인 화면

그림 3. 어플리케이션 구동 화면

1. 어플리케이션 구동 화면
2. 초기 화면으로 검색바를 통해 원하는 영화를 검색할 수 있고, 아래 박스오피스 탭을 통해 상위권에 랭크한 최신 영화들을 간편하게 확인할 수 있다.
 - 2-1. 박스오피스에 있는 영화 선택시 화면으로, 영화의 정보 및 줄거리를 확인할 수 있다.
3. 영화 검색 화면으로 검색 결과 나오는 영화 목록을 확인할 수 있다.
4. 검색 결과에서 나온 영화 중 원하는 리뷰를 선택하면 나오는 화면으로 필터링 된 리뷰들을 확인할 수 있다.
5. 감상 포인트 분석 버튼을 누르면 나오는 화면으로 감상 포인트 추출 모델이 분석한 결과를 바 차트 형태로 시각화하여 보여준다.

7. 캡스톤디자인 결과의 활용방안

- 사회적 기대효과

대다수의 소비자들이 소비 후 불만족을 줄이기 위해 온라인 리뷰를 확인한다. 하지만 이를 악용하여 일부로 부정적인 리뷰를 작성하거나 실제로 소비를 하지 않았음에도 한 것처럼 작성한 리뷰들이 많아 실제 소비자들이 리뷰를 확인하는 것에 많은 피로도를 느끼고 있다. 이를 필터링해 소비자들에게 도움이 될만한 리뷰를 보여줌으로써 사용자들에게 좀 더 객관적인 평가와 정확한 사실을 제공할 권리를 보장하고, 또한 만족 포인트를 텍스트 내에서 추출해 요약하여 보여주기 때문에 소비자들의 질 높은 소비 활동이 가능해질 것으로 기대된다.

- 기술적 기대효과

다양한 상품에 대한 리뷰가 넘쳐나는 현재, 리뷰 데이터를 분석해 소비자의 만족도를 끌어올리는 것은 많은 이커머스 기업들의 목표이다. 속성 기반 감성분석은 기존 감성 분석 모델보다 발전된 형태의 감성 분석 기술로 국내에서도 관련 연구가 이루어지고 있지만 현재까지는 적극적으로 활용되고 있지 않다. 본 캡스톤 디자인은 영화 리뷰에만 이를 적용했지만 이에 국한되지 않고, 다양한 소비 영역으로 확장할 수 있을 것이라 생각된다.