Word Cloud

<여기 앞에 무슨 소제목을 지어주면 좋을 것 같은데 어… 잘 모르겠습니다…>

본격적으로 분석을 진행하기에 앞서, 조사와 접사 등 한국어 특성상 어떤 문장에서든 자주 나타나는 형태소를 제거하고 Word Cloud 이미지를 도출해내는 것이 평점별 영화 댓글을 살펴보는 데에 더 효과적이라고 생각했다. 이를 위해서 형태소를 실제적 의미를 담고 있는 실질 형태소와 문법적 역할을 담당하는 형식 형태소로 구분하였고, 이중에서 실질 형태소만을 Word Cloud 이미지 도출에 활용하였다. Word Cloud 이미지 도출에 사용된 품사 목록은 트위터 태거를 기준으로 Adjective, Adverb, Alpha, Exclamation, Foreign, Hashtag, Korean Particle, Modifier, Noun, Number, Verb의 총 11가지이다.

excel의 필터 기능을 사용하여 위에서 제시한 11가지 품사를 선별한 뒤로는, 각 영화별 리뷰 댓글을 다시 평점에 따라 1점부터 3점까지, 4점부터 7점까지, 8점부터 10점까지의 3단계로 나누어 https://www.jasondavies.com/wordcloud/의 Word Cloud Generator를 통해 Word Cloud 이미지를 도출하였다. 해당 Word Cloud 이미지 도출 전 설정한 설정값은 Spiral: Archimedean, Scale: √n, 5 orientations from 0° to 0°, Numbers of words: 250이다.

<Word Cloud 분석 결과 각 분류 별 자주 등장하는 단어>

• 변성현 감독의 “불한당”

– 모든 평점에서 전반적으로: 영화 / 배우 등

– 평점 1점부터 3점까지: 일베 / 홍어 / 전라도 / sns / 트위터 / 인성 / 새끼 / 망한 / 쓰레기 / 1 / 0 / 재미없어요 / 재미없음 / 입니다 / 마세요 / 합니다 / 드립니다 / 아깝다 / 없다 / ㅉㅉㅉ / ㅉㅉ / ㅠㅠ / 휴 / ㅋㅋㅋ 등 → 의성어나 감정을 표현하는 한글 자모로 많은 것을 설명하기도 하고, 숫자 ‘1’과 ‘0’은 “1점도 아깝다”나 “0점은 왜 안 만들어 주시나요” 등에 사용되었다.

– 평점 4점부터 7점까지: 신세계 / 좋은 / 좋았다 / 논란 / 그냥 / 킬링타임 / 그럭저럭 / 짬뽕 / 아쉽다 / 수준 / 솔직히 / 비슷한 / 괜찮은 / 그저 / 캐스팅 / 브로맨스 / 조폭 / 김희원 / 깡패 등

– 평점 8점부터 10점까지: 느와르 / 연기 / 스토리 / 연기력 / 좋았습니다 / 였습니다 / 봤습니다 / 였다 / 재밌었어요 / 재밌어요 / 좋았어요 / 보세요 / 봤어요 / 합니다 / 좋았다 등

– 모든 평점에서 전반적으로 많이 등장한 ‘감독’ 키워드는 낮은 평점에서보다 높은 평점에서 더 작게 나타나고, ‘임시완’, ‘설경구’ 키워드는 높은 평점으로 갈수록 더 크게 나타나는 경향이 있다.

– ‘신세계’, ‘프리즌’, ‘무간도’ 등 여타 느와르 영화가 비교 대상으로 댓글에 등장하지만 댓글에서 말하고자 하는 바는 평점에 따라 완전히 반대된다. 평점이 높은 쪽에서는 “~ 영화 보다 낫다!” 등으로 언급되지만, 평점이 낮은 쪽에서는 “감히 ~랑 비교를 해?” 등으로 언급된다.

• 홍상수 감독의 “지금은맞고그때는틀리다”, “밤의 해변에서 혼자”

– 모든 평점에서 전반적으로: 영화 / 감독 / 사랑 / 홍상수 / 김민희 등 → 특히 ‘홍상수’ ‘김민희’는 평점이 높은 쪽에서 약간 더 크게 나타난다.

– 평점 1점부터 3점까지: 불륜 / 쓰레기 / 아깝다 / 미화 / 합리화 / 최악 / 더럽다 / 역겹다 / 싫다 / 같다 / ㅉㅉㅉ / 간통죄 / 자식 / 부인 / 유부 / 딸 / 더러워 / 바람 / 열자 등 → ‘불륜’은 평점이 낮은 댓글에서 많이 등장한 편이고, ‘열자’의 경우, 댓글 작성의 최소 기준인 10자를 쓰기도 귀찮다거나 시간이 아깝다며 반복적으로 사용하는 경우가 있다.

– 평점 4점부터 7점까지: 그래도 / 좋았다 / 조금 / 지루한 / 재밌게 / 변명 / 매력 / 틀리다 /사실 / 찌질 / 해변 등

– 평점 8점부터 10점까지: 연기 / 정재영 / 배우 / 좋았다 / 최고 / 대사 / 장면 / 재미 / 일상 / 주인공 / 감정 / 좋다 / 봤습니다 / 좋았어요 / 합니다 / 봤어요 / 입니다 / 좋아요 / 같아요 등

• 니키 카로 감독의 “뮬란(2020)”

– 모든 평점에서 전반적으로: 영화 / 디즈니 / 뮬란 / 원작 / 스토리 / 스포일러 등

– 평점 1점부터 3점까지: 중국 / 홍콩 / 평점 / free / hongkong / 위구르 / 공산당 / 중국영화 /중공 / 신장 / 조선족 / 인권 / 짱개 / 자본 / 쓰레기 / 보이콧 / 이딴 / 알바 / 최악 / 실망 / 2 / 3 / 아니다 / 아깝다 / 없음 / 봐라 / 수준 등 → 2D, 3D와 같이 원작과의 비교를 위해 사용된 것으로 보이고, ‘수준’과 같이 중립적인 단어가 혹평을 위해 사용되었다.

– 평점 4점부터 7점까지: 좋아하는 / 아쉬움 / 솔직히 / 그저 / 논란 / 나름 / 그나마 / 그냥 / 자체 / 수준 / 중국무협 / 액션 / 견자단 / 공리 / 실사영화 등

– 평점 8점부터 10점까지: 유역비 / 액션 / 감동 / 감상 / 입니다 / 재미있게 / 좋은 / 좋았어요 / 좋았음 / 좋아하는 / 봤습니다 / 재밌었어요 / 봤네요 / 였습니다 / 보세요 / 좋았습니다 / 같아요 등

→ 기본적으로 평점이 낮은 댓글에서 부정적 이미지의 단어가 많은 편이고, 전반적으로 영화 자체(스토리나 연기 등)에 대한 내용보다는 스캔들과 관련한 댓글이 많이 보이는 편이다. 반대로 평점이 높은 평점에서는 긍정적 이미지의 단어가 많고, 스캔들 자체의 언급보다 영화의 내부적 내용, 즉 배우나 연기 등에 초점이 맞춰져 있다.

→ 또 평점이 낮은 댓글에 비해 평점이 높은 댓글에서 비해 전반적으로 ‘~니다’ 나 ‘~요’의 어미가 많이 등장해서, 비교적 더 공손하고 정제된 말투를 보여준다.

→ “뮬란(2020)” 경우, 모든 평점 구간에서 ‘스포일러’가 꽤 큰 비중을 차지했는데, 이는 스포일러 내용이 댓글에 포함되면 자동으로 내용이 “스포일러가 포함된 감상평입니다. 감상평 보기”로 바뀌기 때문인 것으로 보인다.

→ 평점 4점부터 7점까지의 댓글은 세 영화의 전반적인 키워드가 비슷한 편이고, 평점 1점부터 3점까지의 댓글과 8점부터 10점까지의 댓글에 등장하는 키워드가 겹쳐서 나타나는 경우가 많다. 평점 1점부터 3점까지의 댓글에는 영화 내용보다는 논란에 집중하고, 평점 8점부터 10점까지의 댓글에서도 “리뷰 칸에서는 영화 이야기만 해라!!”나 “논란은 아쉽지만~” 정도만 언급하고 넘어가는 반면, 평점 4점부터 7점까지의 댓글의 키워드가 영화 자체에 집중하며, 가장 순수하게 영화 내용에 대해서만 말하고 있다는 생각이 들었다.