

빅데이터분석응용실습 기말고사 과제

- “BTS” 키워드로 워드클라우드 출력 및 분석하기 -

영어영문학과(휴먼ICT) 201614010 박도준



위 그림은 “BTS” 키워드로 검색하여 얻은 네이버 뉴스의 기사 중 명사의 출현 빈도를 분석하여 얻은 워드클라우드 이미지이다.

출현 단어 빈도 분석을 위한 선행 작업으로 명사를 추출 과정에서 형태소 분석기인 “RmecabKo” 라이브러리를 사용하였는데, 위 형태소분석에서는 검색 키워드였던 BTS는 명사로 분류하지 않아 위 그림에는 포함되지 않았다. 직접 “BTS” 키워드를 형태소 분석 이전 텍스트에서 확인해본 결과 총 65번의 출현 빈도를 보이며 두 번째로 높은 빈도 순위를 보여주었다.

검색 키워드였던 “BTS”를 제치고 가장 높은 출현 빈도를 보인 단어는 “빌보드”(77회)였다. BTS는 지난 주 발매한 앨범 [BE]와 함께 타이틀 곡 “Life Goes On”을 발표하였다. 빌보드 메인 차트 중에서도 가장 영향력이 있는 Hot100 차트는 일주일 기준 음원 판매량, 미국 라디오 방송 횟수 등을 반영하여 순위를 발표하는데, 지난 달 싱글 앨범으로 발표하였던 “Dynamite”가 Hot100에 오른 이후, 영어가 아닌 한국어 곡이 최초로 Hot100에 오를 수 있을 것에 대한 기대감이 크게 반영된 것이라는 분석이다.

다음으로 눈에 띄는 단어는 "방탄소년단"이다. "방탄소년단"(53회)는 BTS의 한국어 이름으로 BTS와 함께 묶여 언급되었다. 이례적이라고 평가할 만한 부분은, BTS는 한국 남성 그룹 가수로서 처음부터 방탄소년단이라는 그룹 이름을 사용하였음에도 현재 검색되는 뉴스 기사에서는 "방탄소년단"(53회) 보다는 "BTS"(65회)의 출현 빈도가 더 높다는 점이다. 이는 BTS에 대한 현재의 관심이 해외에서 어떠한 평가를 받는 지에 맞춰져 있기 때문인 것으로 볼 수 있다. 뿐만 아니라 BTS의 해외 인기가 상승함에 따라 국내에서도 방탄소년단이라는 이름 보다는 BTS라는 이름을 더 자주 사용하게 된 현상도 이유로 판단된다.

"방탄소년단"(53회)에 뒤를 이어 "앨범"(51회)에 대한 언급도 높음을 볼 수 있다. 이는 앨범 판매량 기준으로 순위를 집계하는 빌보드 Hot200 차트에서 이미 1위가 발표된 영향이라 분석되며, 앨범에 대한 언급보다 빌보드의 단어 출현이 더 높은 점은 현재 언론의 관심이 BTS의 앨범 출시 보다는 빌보드에서의 차트 순위에 더 높은 관심이 몰려 있다고 볼 수 있는 부분이다.

"미국"(22회), "현지"(16회) 키워드와 함께 차트(28회), 정상(16회), 글로벌(6회)이라는 단어도 눈에 띄는데, 이 역시 빌보드와 관련이 있는 단어로서 BTS의 빌보드 순위에 대한 기대감과 현재 전세계적으로 상승하고 있는 인기를 대변하는 부분이라 설명할 수 있다.

그외에 빈도수는 높지 않지만 워드클라우드에 함께 포함된 단어로는 "롯데"(4회), "백화점"(4회), "기생충"(1회), "봉준호"(1회), "코로나"(1회) 등이 있다. 이는 현재 롯데백화점에서 진행 중인 BTS의 팝업스토어, 한국의 위상을 세계에 알린 봉준호 감독의 기생충, 현재의 코로나 시국의 영향이 반영되어 함께 언급 된 것으로 사료된다.