빅데이터분석응용실습 HW9

- "여행" 키워드로 워드클라우드 출력 및 분석하기 -

영어영문학과(휴먼ICT) 201614010 박도준

위 그림은 네이버 블로그에서 "여행"이라는 키워드를 검색하여 나온 포스트에서 단어 출현 빈도를 분석하여 출력한 워드클라우드 그림이다.

위 출력 결과에서는 검색 키워드였던 "여행"을 제외하였는데, 이는 다른 단어들이 1~5번 출현한 것에 비해 "여행"은 41번으로 압도적으로 많이 출현하여, 이를 결과물에 포함시켰을 때 다른 단어들 간의 상대적인 차이를 시각적으로 파악하는데 어려움이 있었기 때문이다.

워드클라우드 그림에서는 단어의 출현 빈도에 따라 폰트의 크기가 달라지며, 빈도 수가 낮아짐에 따라 단어의 위치가 원의 중심에서부터 가장자리로 이동하게 된다(기본 설정의 경우). 위 그림에서는 가장 중심에 위치한 단어가 "제주"였는데, "제주도" 키워드를 포함하였을 경우 총 8번(각각 5번, 3번) 등장하였다. 이는 검색일(11월 4일) 기준 가장 많은 여행객이 다녀간 여행지가 제주도일 것으로 추정해 볼 수 있는 대목이다.

다음으로는 여수(4번), 거제(3번), 중국(3번), 바다(3번) 등과 같은 여행지 단어와 함께 숙소(3번), 코스(3번), 할인(3번), 민박(2번)과 같은 단어들이 함께 눈에 들어온다. 이를 통해 여행지에 대한 언급과 더불어 숙박시설이나 여행 정보에 대해서도 함께 언급되고 있음을 추정해 볼 수 있다.

중국을 제외한 해외 여행지에 대한 언급도 있었는데, 독일, 스페인, 코타키나블루가 각각 한 번씩 등장하였다. 이는 앞서 언급되었던 국내 여행지와 비교해보았을 때 다소 낮은 빈도이다. 이를 통해 현재의 여행 트렌드가 외국이 아닌 국내에 맞춰져 있다는 해석이 가능하다.