

인천공항 트렌드 분석 플랫폼을 이용한 2018년 인천공항 트렌드 분석

손석현*, 최유진*

*인천국제공항공사 공항연구소

e-mail: shson@airport.kr, cyj@airport.kr

Analysis of Trends in Incheon Airport using IIAC Trend Analysis Platform

Seokhyun Son*, Yu-Jin Choi*

*Airport Research Institute, Incheon International Airport Corporation

요 약

인천공항은 2001년 개항이후, 비약적으로 발전하고 있으며, 2018년 약 6500만 명의 여객을 수용하는 등 초대형 메가허브공항으로 도약 중이다. 2018년 인천공항은 제2여객터미널을 개장하며, 인천공항에 대한 언론과 대중의 관심을 증폭시켰고, 평창동계올림픽 개최와 남북정상회담을 통해 대한민국 중추공항으로서의 이미지를 각인시켰다. 인천국제공항공사는 인천공항의 건설 및 관리, 운영, 주변지역개발, 부대사업 및 기타 국가위탁사업을 하는 국토교통부 산하 공공기관으로 2018년 입국장 면세점 도입, 정규직 전환 등의 이슈를 생성하였다. 본 논문에서는 뉴스, SNS 수집, 저장, 처리, 분석플랫폼인 ITAP를 이용하여 인천공항, 인천국제공항공사 관련된 뉴스, SNS 내용을 플랫폼별로 분석하였고, 월별 버즈량과 주요키워드를 플랫폼별로 추출하여 2018년 인천공항, 인천국제공항공사의 주요 이슈를 제시한다.

1. 서론

최근 언론보도에 따르면, 2018년 인천공항의 공항평가지수는 전세계 5위로 2017년 7위 대비 2단계 상승하였다. 2018년 인천공항은 제2여객터미널 개장, 입국장 면세점 도입 등 여객편의를 위한 시설을 더욱 강화하였고, 평창동계올림픽과 남북정상회담 등의 국제적 이벤트는 전세계적으로 인천공항의 대한민국 관문이미지를 더욱 각인시켰다. 인천국제공항공사는 인천공항의 효율적인 운영과 부대사업을 수행하는 국토교통부 산하기관으로 2018년 정규직전환 등의 이슈가 대중과 언론으로부터 주목 받았다.

공공기관을 비롯하여 민간사기업 등 기업 및 기관은 대중의 관심사항 파악, 이미지제고, 신규 서비스개발 등을 위해 뉴스, 사회관계망서비스(SNS)의 언급을 분석하고 모니터링을 하고 있는데, 분석을 위한 방법론으로는 버즈량(언급량) 분석, 주요키워드 추출, 감성분석 등이 있다.

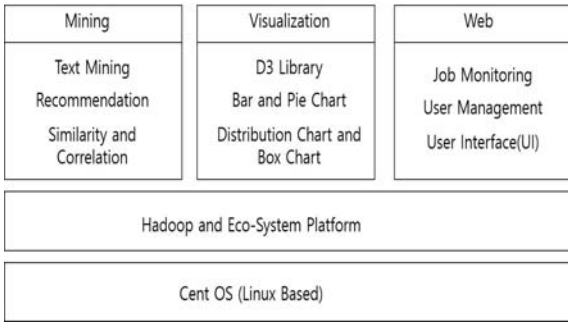
문정훈 외는 빅데이터 분석기술을 이용하여 SNS언급되는 맛집정보와 버즈량을 분석하여 농식품 추천시스템을 연구하였고[1], 한국보건사회연구원에서는 사회복지와 관련된 뉴스, SNS 버즈량을 통해 월별 사회복지 이슈를 분석하고 시사점을 제시하였다[2]. 윤유동 외 기타는 SNS 데이터를 수집하고 수집된 데이터를 긍정과 부정으로 나누어 분석, 빅데이터에 대한 국민적 인식을 연구하였다[3].

본 논문에서는 뉴스, SNS 데이터의 수집과 처리, 분석기능을 제공하는 플랫폼인 인천공항 트렌드 분석 플랫폼

(ITAP, Incheon International Airport Trend Analysis Platform)의 구성을 소개하고 ITAP를 이용하여 인천공항과 인천국제공항공사의 2018년 월별 버즈량과 주요키워드를 추출하여 제시한다.

2. 인천공항 트렌드 분석 플랫폼(ITAP)

인천공항 트렌드 분석 플랫폼(ITAP)은 오픈소스인 하둡(Hadoop)을 기반으로 개발되었으며, 키워드를 기반으로 하여 약 50개 언론사의 네이버뉴스와 네이버블로그의 기사(포스트)를 크롤러(Crawler)를 통해 주기적으로 수집한다. 수집된 기사는 텍스트마이닝을 통해 처리되고 데이터베이스에 저장되며, TF-IDF, LDA, KNN, K-Means, EM, Opinion Mining 등 데이터마이닝 분석기술을 통해 주요 키워드 추출, 키워드간 관계(유사/상관)분석, 감성분석 등의 분석기능을 제공한다. 분석된 데이터는 Hive를 통해 최종적으로 사용자인터페이스에 표출되며, ITAP의 차트와 시각화는 직관적인 분석을 위해 D3 라이브러리를 사용하였다. 또한, 분석결과를 활용한 추천 엔진이 포함되어 있다. ITAP는 인천국제공항의 트렌드 분석 외에도 수집된 대용량의 외부데이터를 토대로 항공산업 트렌드를 분석하고, 수요예측 서비스와 연계하여 항공수요예측에 영향을 미치는 사회적 요인들을 반영한다. 현재 ITAP는 약 200TB의 정형/비정형 데이터를 저장하고 있으며, 향후, 정밀한 분석 및 예측을 위해 확장할 예정이다.



<그림 1> ITAP 구조

3. 2018 인천공항 트렌드 분석

ITAP 크롤러를 통해, 2018년 1월부터 12월까지 ‘인천공항’과 ‘인천국제공항공사’에 대해 언급된 총 227,496건의 뉴스와 SNS(블로그) 기사를 수집하였는데, 수집결과 뉴스(47,913건)에 비해 블로그(179,583건)에서 많은 언급이 나타났다. 상세 수집데이터는 표 1과 같다.

<표 1> 수집데이터

검색어	뉴스	블로그	합계
인천국제공항공사	6,945	3,508	10,453
인천공항	40,968	176,075	217,043
합계	47,913	179,583	227,496

그림 2는 2018년 ‘인천공항’ 키워드의 월별 온라인 버즈량 추이다. 2018년 중 ‘인천공항’이 가장 많이 언급된 월은 2월로 나타났다. 2018년 2월은 평창올림픽이 개최한 달로 북한평창올림픽 대표팀이 인천공항을 통해 방남하였고 그로인한 뉴스기사가 많이 생성되었다. 두 번째로 많은 버즈량을 나타냈던 1월은 인천공항의 제2여객터미널이 개장하였으며, 이와 관련하여 다수의 기사가 생성되었는데, 이는 제2여객터미널에 대한 국민과 언론의 기대와 관심이 매우 높았다는 것을 의미한다. 이밖에 2018년 3월에 발생한 인천공항 기내식 제조시설 사고 역시 다수의 기사가 생성되었음을 알 수 있다.



<그림 2> ‘인천공항’ 월별 버즈량

그림 3은 온라인 플랫폼(뉴스, 블로그)별 버즈량과 주요 키워드를 제시한다. 주요키워드는 ‘인천공항’, ‘인천국제공항공사’가 언급된 뉴스, 블로그 기사 중 TF-IDF를 통해 단어의 가중치를 계산하였고, 토픽모델 중 하나인 LDA를 이용하여 토픽을 분류[4], 최종적으로는 전문가 분석을 통해 대표 토픽을 정의하였다. 온라인 플랫폼별로 살펴보면, 전반적으로 블로그에서는 9월에 가장 많은 버즈량이 나타났으며, 뉴스는 1월에 가장 높은 버즈량을 나타냈다. 9월 블로그의 주요키워드는 ‘추석’, ‘연휴’, ‘메르스’, ‘면접’ 등으로 나타났으며, 1월 뉴스의 주요키워드로는 ‘터미널’, ‘개장’, ‘대한항공’ 등으로 나타났다. 9월 인천공항과 관련된 블로그에서는 추석연휴를 맞아 인천공항을 방문한 여객이 음식점, 휴게시설 등 인천공항의 시설물을 접하고 이에 대한 포스트를 많이 게재하였고, 인천국제공항공사와 관련된 블로그에서는 면접을 위한 분석과 준비에 대한 글이 많이 게재되었다. 1월 인천공항과 인천국제공항공사 뉴스와 관련된 주요키워드인 ‘터미널’, ‘개장’, ‘대한항공’은 인천공항의 제2여객터미널과 관련된 키워드이다.



<그림 3> 온라인 플랫폼별 버즈량

주요키워드별로 살펴보면, 인천공항 뉴스 기사에서는 ‘제2여객터미널’, ‘대표’, ‘정부’, ‘서비스’의 빈도가 가장 높았으며, 인천국제공항공사는 ‘제2여객터미널’, ‘사장’, ‘면접’이 높은 빈도수를 나타냈다. 인천공항, 인천국제공항공사 블로그에서는 ‘여행’과 ‘채용’이 각각 높은 빈도수를 나타냈는데, 이는, 대중을 대변하는 블로그에서는 여행을 위해 방문하는 인천공항과 인천국제공항공사의 채용에 대해 많은 관심이 있다는 것을 알 수 있다. 뉴스의 주요키워드를 종합해보면, 언론은 2018년 인천공항의 제2여객터미널의 개장과 서비스에 대해 높은 관심을 나타낸 것을 알 수 있다.

4. 결론

본 논문에서는 뉴스, SNS 데이터 수집 플랫폼인 ITAP의 구성과 ITAP를 통해 수집된 약 20만 건의 기사를 분석하여 2018년 인천공항과 인천국제공항공사의 온라인

인 플랫폼별 월별 버즈량과 주요키워드를 분석하였다. ITAP는 오픈소스인 하둡을 기반으로 개발된 플랫폼으로 대용량 정형/비정형 데이터의 수집과 분석이 가능하다. ITAP를 통해, 2018년 인천공항, 인천국제공항공사와 관련된 월별 버즈량, 주요키워드 등을 분석한 결과, ‘제2여객터미널 개장’, ‘면세점’, ‘채용’ 등이 언론과 대중으로부터 많은 관심을 받은 것으로 알 수 있다.

참고문헌

- [1] 문정훈, 장익훈, 최영찬, 김진교, 박진 “빅데이터 기반 소비자 유형별 농식품 추천 시스템 구축 사례” 한국통신학회논문지, 제40권 제5호, 903-913, 2015
- [2] 송태민, 진달래 “2015년 소셜 빅데이터기반 보건복지 이슈 동향분석” 한국보건사회연구원 연구보고서, 2015
- [3] 윤유동, 조재춘, 허윤아, 임희석 “소셜미디어 데이터 분석을 활용한 빅데이터에 대한 인식 변화 비교 분석” 정보처리학회논문지, 제6권 제7호, 371-378, 2017
- [4] D. M. Blei, A. Y. Ng, and M. I. Jordan “Latent Dirichlet allocation” the Journal of machine Learning research, vol. 3, pp. 993-1022, 2003