

식품안전 관련 트위터 정보의 연관 관계 분석 및 시각화

소현수⁰, 강승식, 오세욱
국민대학교 컴퓨터공학부, 국민대학교 식품영양학과

hyss4412@naver.com, sskang@kookmin.ac.kr, swoh@kookmin.ac.kr

Cocurrence Relation Analysis and Visualization in Tweet for Food Safety Domain

Hyun-Su So⁰, Seung-Shik Kang, Se-Wook Oh
School of Computer Science & Foods and Nutrition, Kookmin University

요약

식품안전 사고가 발생했을 때 뉴스, 인터넷 기사를 통해 정보를 인지하기 전에 그 음식을 섭취하는 경우가 발생하는 문제점 최소화하기 위하여 실시간 트윗 분석으로 현재 발생한 식품안전 키워드와 어느 지역에서 발생했는지를 신속하게 파악하고, 키워드 연관관계 분석 프로그램을 활용하여 정확한 정보를 추출한다. 이와 더불어, SNS 등 다양한 정보 소스로부터 추출한 정보를 간단명료하게 파악하기 위해서 워드 클라우드 등 데이터 시각화 기법을 활용하여 시각화로 정보를 제공한다. 이 기법은 식품안전 뿐만 아니라 최근 발생한 콜레라 감염 발생과 같은 문제를 해결하기 위한 방법으로 활용될 수 있을 것이다.

주제어: 식품 안전, 트위터, 트윗, 실시간 분석, 데이터 시각화, 워드 클라우드, 키워드 연관관계

1. 서론

SNS는 뉴스보다 빠르게 정보를 전파할 수 있다. 미국에서 이륙하려던 비행기가 활주로를 벗어나는 사고가 발생했을 때, 승객들은 트위터를 활용해서 구조 요청을 했고, 빠르게 구조가 이뤄졌다. 이처럼, SNS는 실시간 정보를 빠르게 전파할 수 있는 매개체이다. 본 논문은 SNS 중 트윗 정보를 분석하여 실시간으로 식품안전 사고를 인지하고 최대한 빠른 시간내에 조치할 수 있도록 하거나 식품안전 관련 지식정보 자료로서 온톨로지를 구축하는데 도움이 되는 시스템을 구축하고자 한다[1,2].¹⁾

2. 실시간 트위터 분석

트위터 정보를 실시간으로 분석하는 프로그램은 Python에서 제공하는 twython library를 이용하여 구현하였는데 이 프로그램은 실시간으로 올라오는 SNS의 게시물을 활용하여 식품 안전정보를 보다 빠르게 사용자에게 전달하기 위한 것이다[3,4]. 이 프로그램은 2가지 방법으로 정보를 얻을 수 있다. 첫째, 특정 키워드가 포함된 게시물을 검색하여 관련 정보를 사용자에게 제공한다. 예를 들어, '식중독'이라는 키워드를 사용할 경우에 트위터에 식중독이 포함된 트윗 게시물을 실시간으로 전송받는다. 이를 통해, 실시간으로 식중독이 발생하고 있다는 것을 파악할 수 있으며 추가로 게시물 내용을 분석하여 발원지와 발생의 원인 등을 분석한다.

둘째, 특정 GPS 좌표의 주변에서 작성된 게시물 정보를 사용자에게 제공한다. 예를 들어, 국민대학교의 GPS 좌표를 입력해 두면 지정한 영역에서 작성된 게시물을

받을 수 있다. 이를 통해, 실시간으로 해당 지역에서 작성된 게시물들을 분석하여 식품안전과 관련된 정보가 발견되면 이를 사용자에게 실시간으로 제공한다.

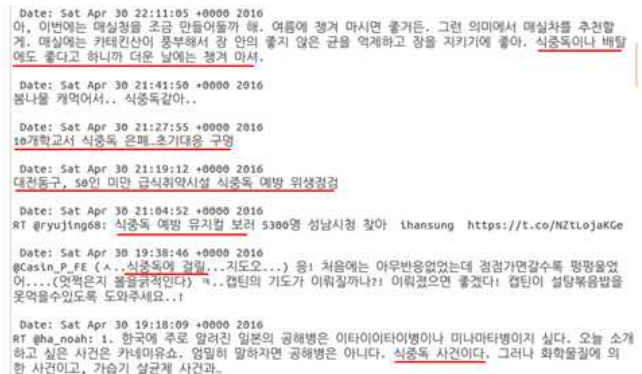


그림 1. 특정 키워드가 포함된 게시물 검색
'식중독' 관련 트윗 정보 예제

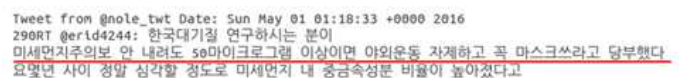


그림 2. 특정 GPS좌표의 주변에서 작성된 게시물 검색
서울특별시 광화문 반경 20Km 정보



그림 3. 특정 GPS좌표의 주변에서 작성된 게시물 검색
인천광역시 주안역 반경 20Km 정보

이 프로그램을 활용하여 얻은 정보는 텍스트 형태가 아니라 그 내용을 시각화하여 사용자에게 한눈에 보여주

1) 이 논문은 2015년 정부(미래창조과학부)의 재원으로 국가과학기술연구회 융합연구단 사업(No. CRC-15-05-ETRI)의 지원을 받아 수행된 연구임.

