

## 공공데이터 및 소셜 웹 데이터를 이용한

### 뷰티 트렌드 분석 서비스

송재오<sup>0</sup>, 김경배<sup>\*</sup>, 이상문<sup>\*\*</sup>

<sup>0</sup>주케이아이씨티 기술연구소

<sup>\*</sup>서원대학교 컴퓨터교육학과

<sup>\*\*</sup>한국교통대학교 컴퓨터정보공학과

e-mail: jos@k-ict.co.kr<sup>0</sup>, gbkim@seowon.ac.kr<sup>\*</sup>, smlee@ut.ac.kr<sup>\*\*</sup>

## Beauty Trend Analysis Services using Public Data and Social Web Data

Je-O Song<sup>0</sup>, Gyoung-Bae Kim<sup>\*</sup>, Sang-Moon Lee<sup>\*\*</sup>

<sup>0</sup>R&D Institute, K-ICT Co.,Ltd.

<sup>\*</sup>Dept. of Computer Education, Seowon University

<sup>\*\*</sup>Dept. of Computer Sci. & Info. Eng., Korea Nat'l Univ. of Transportation

### ● 요약 ●

요우커를 중심으로 한 한류 열풍은 연예인, 미디어 콘텐츠 시장을 넘어서 한국산 제품에도 큰 영향을 미치고 있다. 특히, 화장품 을 비롯한 뷰티 관련 제품은 가장 대표적인 시장으로 주목 받고 있다. 본 논문에서는 공공데이터 및 소셜 웹 데이터를 이용하여 화장품 관련 기업의 비즈니스를 위한 뷰티 트렌드 분석 서비스를 제안한다. 소셜 웹 데이터는 국내의 뷰티 시장에 대한 제품 인지도를 중심으로 분석되며, 공공데이터는 국내에서 유통되는 화장품에 대한 안정성을 중심으로 분석한다.

**키워드:** 공공데이터(public data), 소셜데이터(social data), 웹데이터(web data), 뷰티(beauty), 화장품(cosmetics)

## I. Introduction

행정자치부에서는 국민의 삶의 질을 향상시키고, 지식서비스를 기반으로 하는 창업 생태계를 조성하기 위해 33개 분야 국가중점데이터를 개방하고, 신규 데이터를 국민, 기업, 전문가 수요에 따라 추가로 개방 중에 있다. 특히, 국내 제조업은 세계적인 경제위기 속에 많은 문제와 한계에 직면하고 있다. 이러한 난항을 타개하는 방법으로 제조기술과 ICT기술을 융합하여 새로운 형태 또는 진보된 형태의 융합산업에 대한 요구가 커지고 있다. 이에 본 논문에서는 소셜 웹 빅데이터와 공공데이터의 융합 활용을 위한 데이터 수집/저장 및 분석 리포팅 기술 서비스에 대한 설계 및 구현에 대한 방법을 제안한다.

## II. Implementation and Experiments

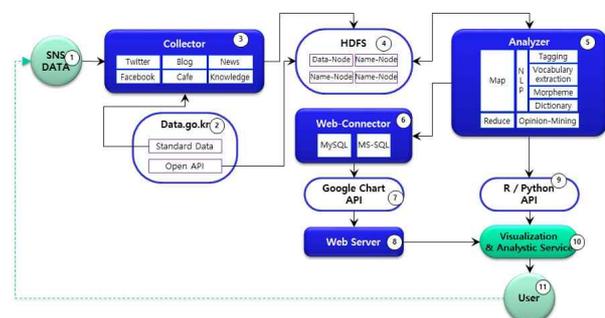


Fig.1. 시스템 구성도

소셜 및 웹 상의 화장품 데이터에 대한 수집 기준은 화장품법시행규칙에 따른 화장품 유형별 키워드를 중심으로 수행되고 그 대상은 트위터, 페이스북, 네이버 카페, 네이버 뉴스, 네이버 지식, 네이버 블로그, 다음 카페, 다음 뉴스 등으로 하고 있다. 공공데이터에서의 화장품 데이터는 한국보건산업진흥원의 화장품 시장규모 및 추이, 기업별 매출동향, 화장품 유형별 생산 현황/추이, FTA에 따른 화장품 교역현황, 모바일/온라인 쇼핑을 통한 화장품 유형별 거래액, 연령대별 생산현황을 대상으로 하며, 식품의약품안전처의 화장품 안전성 정보 (알러지 유발 성분), 배합금지성분, 배합한도성분에 대하여 수집한다.

데이터 수집기는 소셜 웹 데이터와 공공데이터에서 CSV, JSON, RDF, XLS, XML 등의 일반적인 형태를 대상으로 동작하며, OPEN API 형태로 제공되는 공공데이터는 HDFS로 직접 연결되어 저장된다.

수집된 데이터는 비정형 데이터로 수집되는 데이터 소스 채널(출처) 별로 1차 카테고리 구성 후 제품유형, 브랜드/기업, 컬러, 기능별 데이터로 2차 카테고리로 구성하여 NoSQL 형태로 저장된다.

자연어 처리는 오픈소스 형태의 국영문 형태소 분석기를 응용하고 보조적인 사전 구축을 통해 이루어지며, 하둡의 빅데이터 분석구조인 맵리듀스를 따른다. 저장된 데이터를 가져와 자연어 분석하여 화장품에 대한 유형, 브랜드/기업, 컬러, 기능에 대한 키워드를 중심으로 언급회수를 카운팅하고 언어구조를 형태소 형태로 분석하여 카운팅된 키워드의 전후 문장에 대하여 긍정과 부정으로 나누고 연관 키워드 분석을 통해 성분, 구입시기, 화장목적, 화장품광고 등에 대한 선호도와 불만 분석을 수행한다. 이때, 수집된 데이터의 출처를 기준으로 화장품에 대한 이야기가 가장 많이 오고가는 채널을 분석하고 브랜드/기업에 대한 긍정/부정 데이터는 특정 기업/브랜드에 대하여 전체 데이터 대비 포지셔닝을 분석한다.

분석결과는 비정형 형태의 데이터 및 분석정보로 현재 일반적으로 사용하는 웹 브라우저를 통한 서비스를 위해서는 정형화된 데이터로 바뀌어야 한다. 따라서, 분석결과를 웹 형태로 서비스 할 수 있도록 서비스 기능(메뉴)에 따라 정형화 데이터로 변경하여 일반적인 RDBMS 형태인 MySQL 또는 MS-SQL에 저장한다.

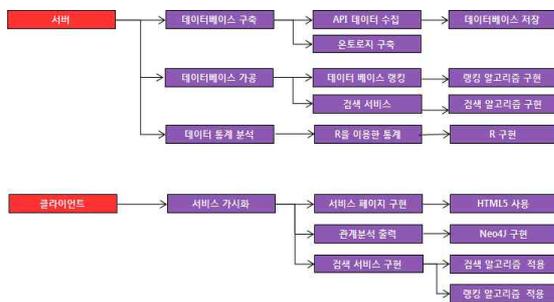


Fig.2. 서비스 구성도

웹 사이트를 통한 서비스는 차트 작성 등을 위한 시각화가 필요하다. Google Chart API를 웹 서비스를 구현하였으며 상황에 따라 Neo4J 및 R을 통해 서비스되는 메뉴도 있다.

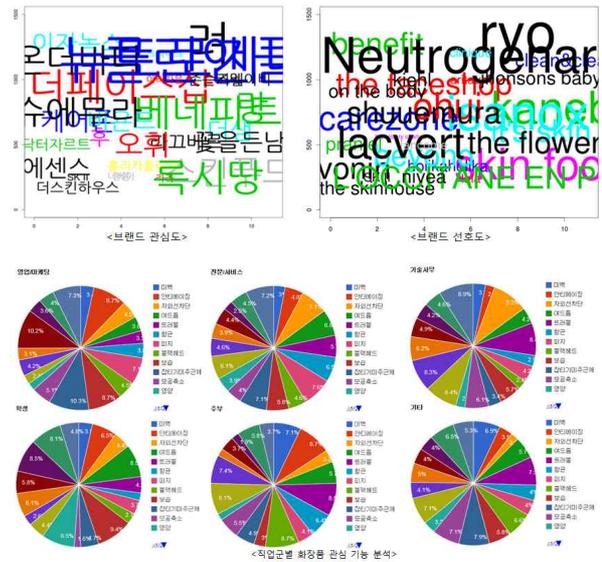


Fig.3. 주요 결과 화면

### III. Conclusions

해당 서비스는 화장품을 중심으로 한 바이오뷰티에 대한 지역특화 산업 관련 기관에서 지역 기업을 위한 컨설팅 도구로 검토되고 있다. 기존의 소셜 웹 데이터 기반 시스템에 공공데이터를 추가하여 분석함으로써, 제품에 대한 불만과 성분에 대한 분석 결과 신뢰성을 높일 수 있었고, 해당 서비스와 연계한 화장품 추천이 가능한 개인 맞춤형 모바일 서비스로 확장이 가능하다.

### Acknowledgments

본 논문은 중소기업청에서 지원하는 2016년도 산학연협력 기술개발사업(No.C0441437)의 연구수행으로 인한 결과물임을 밝힙니다.

### References

- [1] K-ICT CO.,LTD., “System And Method For Analysing Trend Of Cosmetic Custoeem”, Patent, No.10-2016-0178303, 2016.
- [2] Jeo Song, Sang Moon Lee, “A Study on the Application Modeling of SNS Big-data for a Micro-Targeting using K-Means Clustering”, Proceedings of Korea Society of Computer Information, Vol.23, No.1, pp.321-324, 2015.
- [3] Jeo Song, Jin Kwan Kown, Yoon Su Park, Sang Moon Lee, “An Implementation Micro-Targeting Service for Cosmetics using SNS Big-Data Analysis“, Proceeding of the Korea Multimedia Society, Vol.17, No.2, pp.128-131, 2014.