

빅데이터 기반의 상권분석 시스템 구현에 관한 연구

김종원*·박윤보*·류주미*, 신주범**, 박대기***

*금오공과대학교 컴퓨터공학과

**부경대학교 컴퓨터공학과

***전남대학교 수학과

e-mail:gim0903@naver.com

A Study on Implementation of Commercial Analysis System Based on Big Data

Jong-won Kim*, Yoon-bo Park*,

Jo-mi Ryu*, Ju-beom Shin**, Dae-gi Park***

*Dept of Computer Engineering, Kumoh national institute University

**Dept of Computer Engineering, Pukyong National University University

***Chonnam National University

요 약

본 프로젝트의 목적은 소상공인들을 위한 상권 분석, 트렌드 분석, 창업 지원 정책 소개, 커뮤니티 등을 제공하는 빅 데이터 기반의 웹 서비스를 구축하는 것이다. 일반적인 창업 관련 사이트는 정형데이터를 DB(Data Base)에 저장 후 관리되는 시스템으로, 이는 사용자 개개인에 맞는 맞춤형 정보를 제공하기 힘들다. 따라서 본 논문에서는 실시간 검색어 수집 및 분석을 통해 소상공인들이 창업을 희망할 때, 사용자에게 맞는 정보를 제공하는 맞춤형 서비스 연구에 대한 내용이다.

1. 서론

최근 정부 및 각종 공공기관으로부터 추진되는 수많은 창업지원정책으로 취업준비생 등 사람들로 부터 창업에 대한 관심이 높아지고 있다. 현재 창업에 대한 정보를 제공하는 관련 사이트가 여럿 존재한다. 하지만 방대한 양의 정보들은 시간순서대로 나열되어 있을 뿐, 사용자의 상황에 맞는 정보(지역별 정책, 업종별 정보, 유동인구 등)를 찾기엔 다소 많은 시간이 소요된다. 이 정보들을 어떻게 활용하느냐에 따라 창업의 성공 가능성은 좌지우지된다.

따라서 본 프로젝트에서는 기존 실시간 크롤링(웹 정보 수집), 정형데이터 가시화, 데이터 마이닝 등 IT기술을 활용하여 소상공인들에 알맞은 정보들을 빠르게 알려주고, 분석해주는 솔루션을 다룬다.

본 논문의 구성은 다음과 같다. 2장에서는 프로젝트에 쓰인 관련 기술, 알고리즘을 서술한다.

3장에서는 정형·비정형 데이터들을 시각화한 방법에 대해 설명한다. 4장에서는 정제된 데이터들을 보여주는 프레임워크의 구조의 대해 설명한다. 마지막으로 프로젝트 결과와 결론 및 향후과제에 대해 기술한다.

2. 사용 기술 및 알고리즘

2.1 크롤링

컴퓨터 소프트웨어 기술로 웹 사이트들에서 원하는 정

보를 추출하는 것을 의미한다. 엄청난 분량의 웹문서를 사람이 일일이 구별해서 모으는 일은 불가능에 가깝다. 때문에 웹 문서 검색에서는 사람이 일일이 하는 대신 이를 자동으로 수행해줍니다 이 프로젝트에서는 창업에 대한 정책, 지원 등의 공고 및 뉴스, 기사의 헤드라인 등을 크롤링하여 보여준다.



[그림 1] 크롤링의 기본 구조

2.2 빅데이터 분석

2.2.1 데이터 마이닝

전국 상권 데이터, 관련 비정형 데이터들의 각각 변수들을 선정, 분석하여 소상공인들에게 지역경쟁력, 업종 경쟁력, 상권 안정성을 점수로 각각 나타낸다.

지역경쟁력 점수는 사용자들의 희망 지역 시도별·구

별·업종을 입력값으로 하여 수식 (1)와 같이 계산되며, 상대적 점수로 나타낸다.

$$(1 - \frac{\text{희망지역(구)별상가개수}}{\text{희망지역(시도)별상가개수}}) * (\frac{100}{\text{희망지역(구)개수}}) \quad (1)$$

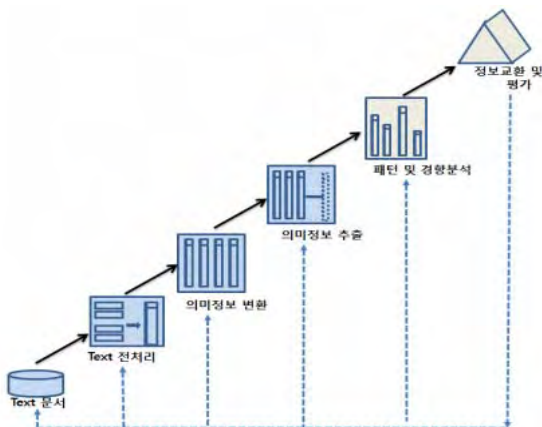
업종경쟁력 점수는 업종별 상가 개수를 정렬하여 절대적 점수로 수치화하는 알고리즘을 사용하였다.

상권안정성 점수는 분기별 건물을 기준으로 하여, 다음 분기까지 남아있는 기간을 점수화 하고 수식 (2)와 같이 계산된다.

$$\frac{\text{희망지역별안정성 점수총합}}{(5 * \text{희망지역별상가갯수})} * 100 \quad (2)$$

2.2.2 텍스트 마이닝

사용자의 희망 업종에 관한 SNS의 실시간 트윗·해시태그 등을 크롤링하고 이때, 긍정·부정·중립에 관련된 단어들이 저장된 감정 사전을 활용하여 크롤링한 SNS 문장들과 비교 및 분석 하여 감정 그래프를 그린다.



[그림 2] 텍스트 마이닝 과정

또한 희망 업종에 관련된 키워드를 모아 순위를 매겨 워드클라우드를 활용해 시각화한다.

3 데이터시각화

3.1 지도 시각화

Google Maps JavaScript API를 이용하여 상권 위치, 유동인구 밀도를 시각화한다.

상권 정보가 저장된 .csv파일에서 좌표정보를 분리하여 JSON파일로 만들어 저장, 로드한다. 로드된 좌표값을 통해 상권별로의 위치를 지도에 뿌려준다.

3.2 그래프 시각화

데이터 시각화 라이브러리인 chart.js, d3.js를 사용하여 상권 위치정보를 그래프로 그려준다. 그림.2에서 추출한 연관성 있는 단어들을 그림.3 으로 보여 줄 수 있다.



[그림 3] 부동산·임대 워드클라우드

[표 1] Chart.js 구성 옵션

Name	Type	Default	Description
responsive	Boolean	true	컨테이너가있을 때 차트 캔버스의 크기를 조정
responsiveAnimationDuration	Number	0	크기 변경 이벤트 후 새 크기로 애니메이션을 적용하는데 걸리는 시간
maintainAspectRatio	Boolean	true	크기를 조정할 때 원본 캔버스 종횡비 유지
onResize	Function	null	크기 변경이 발생할 때 호출됩니다. 차트 인스턴스와 새 크기라는 두 개의 인수를 가져옴.

4. 프레임워크

본 프로젝트는 대시보드를 활용해 정제된 데이터들을 시각화하여 보여준다. 대시보드가 일반적으로 사용하는 PC 환경에서만 모습이 아닌, 소형 디스플레이의 모바일 화면에서도 최적화 된 모습을 보일 수 있도록 Bootstrap을 사용하였다. Bootstrap은 반응형 웹 페이지를 만들 수 있게 도와주는 프레임워크이다.

5. 결과 및 결론

데이터시각화망 지역·업종을 입력하면 그에 대한 정보들이 대시보드 상에 표시된다. 소상공인, 예비 창업인을 위해 만들어진 사이트에서 주기적인 크롤링을 통하여 창업에 필요한 정보·정책들을 보여준다. Google Maps API를 통해 유동인구, 상권 정보·위치를 대시보드 위에 지도로 표시 해주며 지역에 맞는 SNS 이슈 연관키워드를 보여주는 서비스를 제공한다. 대시보드에 접속한 소상공인, 예비창업자 사이의 커뮤니케이션이 가능하다. Twitter API를 통해 실시간으로 SNS에서 사용자가 선택한 직종에 대한 연관 단어를 가져와 직종에 대한 관심도에 따른 워드클라우드를 보여주며, 감정분석한 결과를 점수로 보여준다. 희망 지역이 속한 도시의 희망 업종의 분포, 기간별 추이 등을 Chart.js, D3.js를 활용, 시각화하여 각종 그래프의 형

태로 보여준다. 상권 정보, 유동인구밀도, SNS의 감정·평판·관심도의 정형·비정형 데이터들을 활용하여 데이터마케팅을 거쳐 소상공인들에게 창업의 위험도를 점수로 알려주는 서비스를 구현했다.



[그림 4] 시스템 기능

본 프로젝트를 통해서 빅데이터에 관한 내용에 대해 다양하게 학습할 수 있었다. 빅데이터를 이해하기 위해선 데이터베이스(SQL, NoSQL), 서버 관리, 분석 알고리즘에 대한 깊이 있는 이해가 필요함을 알 수 있었다. 본 프로젝트에서는 상권 정보에 대한 분석으로 한정 하였으나 앞으로의 프로젝트에서는 좀 더 다양한 데이터들을 기반으로 변수들 사이의 관계에 대해 더 공부하고 분석 알고리즘에 대해 공부를 하여 좀 더 효용성과 사용성이 뛰어난 결과물이 나올 수 있도록 심도 있고 세분화된 연구가 필요할 것으로 사료된다.

참고문헌

[1]김정숙, 빅데이터 활용과 관련기술 고찰, 삼육대학교 석사학위논문, 2015

[2]김형년, 빅데이터 비주얼라이제이션 사례분석 -시각화 표현 방식을 중심으로-, 경기대학교, 2014

[3]이재중, 박진완, 웹 데이터를 이용한 이미지의 텍스트 시각화 구현, 중앙대학교, 2011