

빅 데이터 기반의 상권 서비스 확장을 위한 설문조사시스템 설계 및 구현

Design and Implementation of a Survey System for Expanding Big Data-Based Commercial District Service

이원철¹ · 강만수² · 김진호^{1*}

강원대학교 대학원 컴퓨터학과¹, 서울신용보증재단 소상공인정책연구센터²

요 약

우리나라의 영세 소상공인과 자영업자의 비중이 주요 선진국에 비해 과도하게 높고 빈번한 창업과 폐업이 반복되어 국가 경제에 막대한 피해를 초래하고 있다. 이러한 문제를 해결하기 위해 소상공인을 위한 다양한 연구가 진행 중이며, 정부는 소상공인을 위해 빅 데이터를 이용한 상권정보 분석 서비스를 제공하고 있다. 상권정보 분석 서비스 중 서울시에서 운영하는 우리마을가게 상권분석서비스는 소상공인 관련 빅 데이터 분석 서비스를 제공하기 위해 지속적인 서비스 개선을 진행하고 있다. 그러나 다양한 기관에서 제공받은 빅 데이터를 통합하여 서비스를 구축하였기 때문에 데이터 신뢰성의 한계, 데이터 분석의 한계, 서비스 구성의 한계가 존재한다. 이러한 한계를 극복하기 위해 본 논문에서는 빅 데이터 기반의 상권 서비스와 연계 분석이 가능한 위치기반 설문조사시스템을 제안한다. 제안된 설문조사시스템은 설문정보와 상권정보를 연계하여 빅 데이터 상권 분석 서비스를 확장할 수 기반을 마련하였다.

■ 중심어 : 빅 데이터, 소상공인, 설문조사시스템, 연계 분석, 상권정보

Abstract

The proportion of micro-enterprises and self-employed in Korea is excessively high compared to that of major developed countries, and frequent start-ups and business closures are repeated, causing enormous damage to the national economy. In order to solve this problem, various studies are underway for micro-enterprises, and the government provides commercial district information analysis services using big data for micro-enterprises. Among the commercial district information analysis services, the commercial district information analysis of our village store operated by the Seoul Metropolitan Government is continuously improving its service to provide the big data analysis service related to micro-enterprises. Since the service was built by integrating big data provided by various organizations, however, there are limitations in data reliability, data analysis, and service composition. In order to overcome these limitations, this paper proposes a location-based survey system that can be analyzed in conjunction with big data-based commercial district services. The proposed questionnaire survey system established the basis for expending the big data commercial district analysis service by linking the survey information and commercial district information.

■ Keyword : Big data, Micro-enterprises, Survey System, Linkage Analysis, Commercial district information

I. 서론

우리나라의 영세 소상공인과 자영업자의 비중이 주요 선진국에 비해 과도하게 높으며 빈번한 창업과 폐업이 반복되고 있다[1]. 서울의 경우 영세자영업자 통계에서 보여주는 바와 같이 서울시 전체 사업체에서 소상공인은 84.2%를 차지하고 있으며 소상공인 사업체는 지속적으로 증가하고 있다. 이는 진입장벽이 낮은 생계형 창업이 많이 발생하기 때문이다. 그러나 폐업을 또한 매우 높아 사회경제에 막대한 피해를 초래한다. 폐업이 높은 원인은 과다경쟁 및 경기침체에 따른 매출 부진이 가장 많고, 2017년 이후 최저임금 인상 문제와 과도한 임대료 지출도 소상공인에게 어려움을 가중시키고 있다[2].

서울시는 이와 같이 어려움을 겪고 있는 소상공인 및 자영업자에게 도움을 주고자 2016년부터 빅 데이터를 이용한 우리마을가게 상권분석서비스를 구축하여 서비스하고 있다[3]. 상권분석 서비스를 통하여 소상공인영역에 진입하고자 하는 예비 창업자에게 사업관련 정보를 제공하여 폐업률을 낮추고, 자율적으로 무분별한 진입을 자제하고자 하였다.

우리마을가게 상권분석서비스는 소상공인 관련 상권정보의 정량적인 데이터 분석을 통해 소상공인 정책 수립을 지원하고, 소상공인의 창업 정보를 활용하여 국가 및 지역 경제가 안정적으로 지탱할 수 있는 기반을 제시하고 있으며, 영세 자영업자의 경영실태 및 다양한 업종별 특성을 제공하여 국가 정책의 지원방안 수립을 위한 근거 자료로 활용되고 있다.

우리마을가게 상권서비스는 서비스 영역으로 블록, 동, 구, 시를 포함하고 있으며, 생활밀접업종으로 외식업, 소매업, 서비스업의 3개 분야 총 100개의 업종으로 데이터가 구성되어 있다. 그리고 시간과 서비스 영역을 기준으로 업종별 창업 위험도, 서울시내 주요 상권 검색 및 분석, 내 점

포 분석, 지역별 현황 및 상권 동향 등 다양한 빅 데이터 분석 서비스가 제공되고 있다.

그러나 상권분석 서비스 구축을 위해 다양한 기관에서 제공받은 빅 데이터를 통합 및 가공하고 있어 다음과 같은 데이터의 한계점을 가지고 있다. 첫째, 다양한 기관에서 제공받은 빅 데이터를 통합 가공하여 상권 분석 서비스를 만들었기 때문에 데이터에 대한 정합성 검증이 어렵다. 둘째, 제공받는 데이터가 제한되어 있어 상권 데이터 분석의 한계가 있다. 셋째, 기관에서 개별로 식별이 가능한 데이터 형태로 제공되던 데이터가 개인정보보호 등의 사유로 점차 집계데이터 형태로 제공됨에 따라 세부 영역으로 제공되던 서비스가 불가능해지고 있다.

본 논문에서는 이와 같은 데이터 검증, 데이터 분석의 한계, 그리고 집계데이터 제공의 문제를 해결하기 위해 설문결과를 상권정보와 연계하여 분석한다. 설문결과와 상권정보를 연계하는 방법으로 설문결과와 업종과 X, Y 좌표 정보를 이용한 설문조사시스템을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 소상공인을 위한 상권분석 연구와 서비스에 대해서 살펴본다. 그리고 빅 데이터 상권서비스와 연계한 설문조사시스템 구축을 위해 온라인 설문조사 플랫폼에 대해서 설명한다. 3장에서는 빅 데이터 상권서비스와 연계 분석을 위한 설문조사시스템의 구조 및 설계 방법을 설명한다. 4장에서는 구현된 내용을 상권정보 연계를 중심으로 설명하고, 상권정보와 설문결과와의 연계 분석 결과를 보인다. 마지막으로 5장에서는 결론과 향후 연구 방향을 제시한다.

II. 관련 연구

본 절에서는 소상공인을 위한 빅 데이터를 이용한 상권분석 연구와 서비스에 대해서 설명한다. 그리고 빅 데이터 상권서비스와 연계 가능한

온라인 설문 플랫폼 구축을 위해 현재 가장 많이 사용되고 있는 온라인 설문 플랫폼을 소개하고 상권정보와 연계분석을 위한 설문의 위치정보에 대하여 설명한다.

2.1 소상공인을 위한 상권분석 연구

소상공인을 위한 상권분석 연구는 크게 입지 분석과 골목상권 성장요인분석, 상권의 동태적 변화 분석, 그리고 최근에는 소상공인의 마케팅을 돕기 위해 소셜 빅 데이터를 활용한 연구가 진행되었다. 입지분석은 빅 데이터 분석을 통해 적합한 골목상권을 선정하는 방법으로 유동인구, 상점수, 업종별 위험도 등을 분석하여 최적의 입지를 추천해 준다[4]. 골목상권 성장요인 분석은 골목상권의 성장에 영향을 미치는 요인을 찾아 상권의 성장 요인을 분석하는 방법이다[5]. 상권의 동태적 변화 분석은 상권의 유형별로 영향을 미치는 요인을 도출하고 상권의 동태적인 변화과정을 추적하여 상권의 유형별 변화를 파악한다[6]. 그리고 소셜 빅 데이터를 활용한 연구는 업체별로 리뷰되는 소셜 데이터와 지역별, 업종별로 유지되는 요약정보, 그리고 지역별 공공 데이터를 기반으로 소상공인들의 마케팅에 도움이 될 수 있는 지표를 도출한다[7].

2.2 소상공인을 위해 제공되는 빅 데이터 기반 상권 분석 서비스

예비창업자 및 자영업자에게 빅 데이터 기반 분석 상권정보를 제공하는 곳은 소상공인시장진흥공단에서 2006년부터 서비스하는 상권정보시스템[8]과 서울시가 2016년부터 독자적으로 구축하여 소상공인에게 필요한 정보를 제공하는 우리마을가게 상권분석서비스, 그리고 경기도가 2018년부터 구축하여 경기도민에게 제공하는 상권영향분석시스템이 있다[9]. 본 논문에서는 현재 가장 적극적으로 서비스를 개선하고 있는 서울시 우리마을가게 상권분석서비스를 대상으로 상권정보와 연계분석이 가능한 설문시스템을 설계 및 구축 방법을 설명한다.

서울시 우리마을가게 상권분석서비스의 구축 목적은 상권의 정량데이터 분석을 통해 소상공인 정책 수립을 지원하고, 소상공인의 창업정보를 활용하여 국가 및 지역 경제가 안정적으로 지탱할 수 있는 기반을 제시하는 것이다.

이를 위해 우리마을가게 상권분석서비스의 데이터 구성은 위치 기반의 빅 데이터로 구성되어 있고, 영역의 포함 범위를 블록 < 동 < 구 < 시, 상권(골목, 발달, 전통)으로 구성된다. 업종 구분은 외식업(10), 소매업(43), 서비스업(47) 3개 분야 총 100개의 생활밀접 업종으로 분류하여 서비스 되고 있으며, 각 데이터와 분석 서비스는 아래

<표 1> 데이터 구성과 분석서비스 표

데이터	세부데이터	분석
상가업소, 사업체 총조사	점포수, 개폐업률/연차별생존율/평균영업기간/창업위험도, 종업원수	업종분석
카드(BC/KB/신한), 통계청(BR)	성별, 연령대, 시간대, 요일별	매출분석
KT유동인구	건물단위 상존인구, 길단위 상존인구	인구분석
주거인구	성별, 연대대별 주거인구	주거 인구분석
직장인구	성별, 연령대별 직장인구	직장 인구분석
금융비용	대출, 연체료, 이자, 임대시세	금융분석
기타	집객시설, 지하철, 버스정류장, 아파트, 건물, 총가구	지역통계

〈표 2〉 상권서비스와 세부서비스 내역

서비스명	세부서비스
개폐업수(률)	개업수, 폐업수, 개업률, 폐업률
매출현황	평균매출액
소득/가구수	소득분위, 가구수
연차별생존율	1년생존율, 3년생존율, 5년생존율
인구수	유동인수, 주거인구, 직장인구
임대시세	환산임대료,(1층), 환산임대료(그외)
점포수	일반점포수, 프랜차이즈 점포수
평균영업기간	최근10년기준, 최근30년기준

의 표와 같다.

우리마을가게 상권분석서비스의 구성은 시간과 공간을 기준으로 분석 서비스를 제공하고 있다. 시간 기준으로 월, 분기, 년을 사용하고, 공간 기준으로 블록, 행정동, 구, 시, 상권(골목, 발달, 재생, 전통, 관광특구), 위치(특정좌표를 기준으로 100m~1km에 포함된 영역)로 구성되어 있다. 업종은 100개 생활밀접 업종으로 분류하였다. 이를 기반으로 아래와 같은 분석 서비스가 제공되고 있다.

그리고 주요 서비스로 업종별 창업위험도, 서울시내 주요 상권검색 및 분석, 내 점포 분석, 지역별 현황 및 상권동향, 상가업소 통계, 창·폐업률, 인구밀도, 추정매출, 임대정보 등 다양한 상권의 정량적 빅 데이터 분석 서비스를 제공한다.

2.3 상권분석서비스의 한계

우리마을가게 상권분석서비스는 소상공인들이 필요한 다양한 정보를 카드사, 국민건강보험공단 등 다양한 기관에서 빅 데이터를 제공받아 이를 통합 및 가공하여 상권분석 서비스를 제공하고 있어 다음과 같은 한계가 있다.

첫째, 다양한 기관에서 제공 받은 빅 데이터를 통합 가공하여 만든 상권데이터는 정합성 검증이 어렵다. 제공받은 데이터를 교차 체크하여 데이터 검증이 필요하지만, 대부분 특정기관에서 제

공받은 데이터로 교차 검증이 어렵다.

둘째, 데이터 부족에 따른 상권 데이터 분석의 한계이다. 기관에서 수급 받은 데이터는 기관에서 보유한 제한된 데이터로 다양한 상권분석에 충분하지 못하다. 예를 들어, 창업 현황 분석에 필요한 정량데이터로 창업자금, 직원고용형태, 직원 수 등이 있고, 정책결정 지원을 위한 분석 데이터로 소상공인 정책 만족도, 폐업원인, 창업동기, 코로나 영향 등이 있다. 그러나 이러한 데이터를 제공받을 수 있는 기관이 없다.

셋째, 데이터 제공 기관은 블록과 같은 작은 영역의 집계성 데이터로 제공하는 방법에서 점차 행정동과 같이 큰 영역의 집계성 데이터로 제공하여 분석서비스 구성이 블록에서 행정동으로 넓어져 기존의 블록 서비스가 불가능하다. 이는 개인정보 보호정책의 강화에 따른 것으로 블록에 특정 업체가 하나만 존재할 경우 그 업체의 매출액 정보를 추정할 수 있기 때문이다.

위와 같은 문제를 해결하기 위해, 상권 정보와 동일한 지역에 대한 설문문을 통해 데이터 분석에 필요한 데이터를 수집하고, 수집된 데이터를 상권정보와 연계하여 분석하는 서비스가 필요하다.

2.4 온라인 설문조사 플랫폼

설문조사란 표본 집단으로부터 정보를 수집한 결과를 모집단으로 일반화하려는 방법이다. 설문의 구성은 설문에 대한 기본 사항인 주제, 목적, 대상, 설문기간 등과 설문문항으로 구성된다. 설문문항을 구성한 후 면대면, 전화설문, 자기관리 종이 설문, 자기관리 컴퓨터설문(온라인 설문)의 방법으로 설문을 진행한다[10].

우리나라에서 가장 많이 사용하고 있는 온라인 설문 플랫폼으로는 구글폼[11], 네이버폼[12], 그리고 서베이 몽키[13]가 있다. 이러한 서비스들은 모두 무료로 제공되고 있으며, 설문을 작성하고 메일 또는 게시판 등을 이용하여 설문을 실시

〈표 3〉 온라인 설문플랫폼 기능 비교표

기능 항목	네이버	구글	서베이몽키
통합관리	불가능	불가능	불가능
오프라인설문의 온라인 통합관리	불가능	불가능	불가능
중복설문방지	로그인, IP중복체크	로그인	로그인, IP중복체크
결과분석범위	기초통계	기초통계	다양한 분석
위치정보사용	미사용	미사용	미사용

할 수 있다. 설문 결과도 기초통계 뿐만 아니라 간단한 그래프 형식으로 결과를 보여준다. 그리고 유료 서비스 사용 시 설문 작성을 위한 다양한 기능을 제공하고, 서베이 몽키는 다양한 통계 분석 기능까지 제공한다.

<표 3>은 온라인 설문플랫폼의 기능 비교표이다. 통합관리란 하나의 플랫폼에서 계층적인 권한구조를 가지고 통합 관리하는 기능이다. 그러나 기존 플랫폼은 개인들이 사용할 수 있지만, 조직 전체가 통합관리 할 수 있는 체계의 서비스를 제공하지 않는다.

오프라인 설문의 온라인 통합관리 기능은 오프라인 설문 결과를 온라인으로 다시 구성하여 데이터를 저장하는 기능으로 기존의 서비스는 제공되지 않는다. 중복 설문 방지만 동일한 사람이 여러 번 설문을 실행하는 것을 방지하는 방법으로 로그인 또는 IP중복체크 기능을 사용한다.

결과분석 범위란 설문결과를 분석하는 방법이다. 기초통계는 하나의 문항에 대한 보기문항 선택 횟수 및 비율을 계산하고, 평균, 최대값, 최소값, 중간값 등을 계산하는 것이다. 서베이 몽키는 다양한 설문문항간의 분석도 가능하다.

위치정보 사용이란 설문 대상자의 위치정보를 수집하는 방법으로 설문 대상자가 설문시스템에 접속했을 때 설문대상자의 위치정보를 시스템에 제공하는 방법이다. 본 논문에서는 설문대상자인 소상공인의 점포 위치를 사용하여 기존의 빅 데이터 기반의 상권분석 정보의 영역 정보와 연계하여 분석한다.

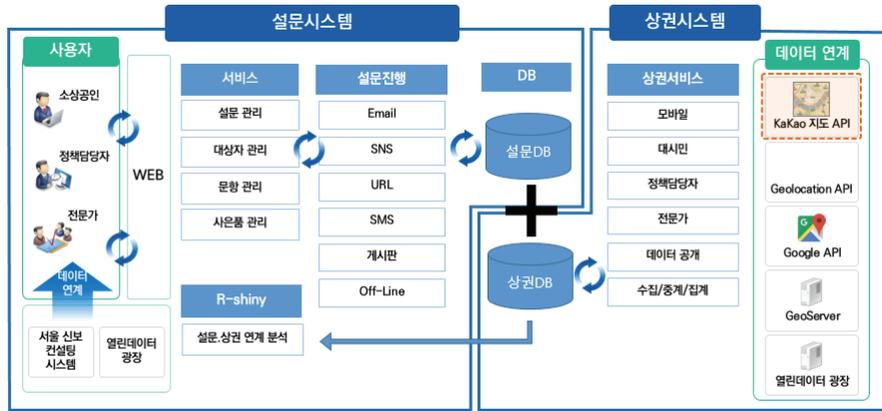
III. 빅 데이터 기반의 상권서비스와 연계 분석을 위한 설문조사시스템 설계

본 논문에서 제안하는 설문조사시스템은 빅 데이터 기반의 상권정보서비스와 연계 분석이 가능한 설문조사 시스템으로 설문작성부터 결과 분석까지 상권분석 서비스와 연계되어 운영이 될 수 있도록 설계된 시스템이다. 이를 위해 첫째, 시스템 구성도를 설명하고, 둘째, 스키마 구조에 대해서 설명한다. 셋째, 각 모듈별 기능에 대해서 설명한다.

3.1 빅 데이터 기반의 상권서비스와 연계 분석을 위한 설문조사시스템 구성도

본 논문에서 제안하는 상권서비스 연계 분석 설문조사시스템의 전체적인 구성도는 그림1과 같다. 시스템의 사용자는 정책담당자, 전문가, 소상공인으로 구성되고 각자의 영역에서 설문시스템을 사용한다. 정책담당자와 전문가는 계층적인 권한 구조로 서비스되고 설문 내용을 공유하지만, 소상공인은 개인화된 서비스로 자신만의 설문을 관리하고 운영한다.

설문서비스는 설문등록, 문항관리, 대상자 관리, 사은품 관리와 설문·상권 연계분석 관리로 구성되어 있다. 설문등록은 설문의 기본정보와 상권에서 연계 분석될 정보를 선택한다. 문항관리를 용이하기 위해 문항은행 관리와 템플릿관리 기능을 추가하여 설문의 일관성을 유지하고, 분



〈그림 1〉 시스템 구성도

석이 용이하도록 한다. 대상자 관리는 패널 그룹을 만들어 설문대상자들에 대한 대표성과 일관성을 갖도록 한다.

설문은 다양한 방법으로 실시가 가능하도록 Mailing 서비스, SMS 연동, SNS 연동, URL서비스, 게시판 공개 설문서비스가 가능하도록 설계하고, Off-Line설문 후 On-Line으로 설문 통합이 가능하도록 구성한다.

DB구성은 설문DB와 상권DB로 분리되어 구성한다. 설문DB는 설문만을 위해 사용하는 DB이고, 상권DB는 기존의 상권서비스가 제공되고 있는 DB이다. 설문이 완료하게 되면 설문DB에 저장된 설문결과와 상권DB에 저장된 상권정보를 좌표와 시점으로 연계하여 R-shiny로 분석한다.

3.2 DB스키마 구조

본 논문에서 제안하는 설문조사 DB스키마는 그림 2와 같다. 설문등록을 시작으로 설문방향 및 보기방향 등록, 설문대상자관리, 그리고 설문결과로 구성된다. 여기에 설문문항을 효율적으로 관리하기 위해 문항은행과 설문 템플릿을 만들고, 설문 사은품과 당첨자를 관리한다. 설문분석 결과는 문항별 설문결과와 개인별 설문결과로 구성한다.

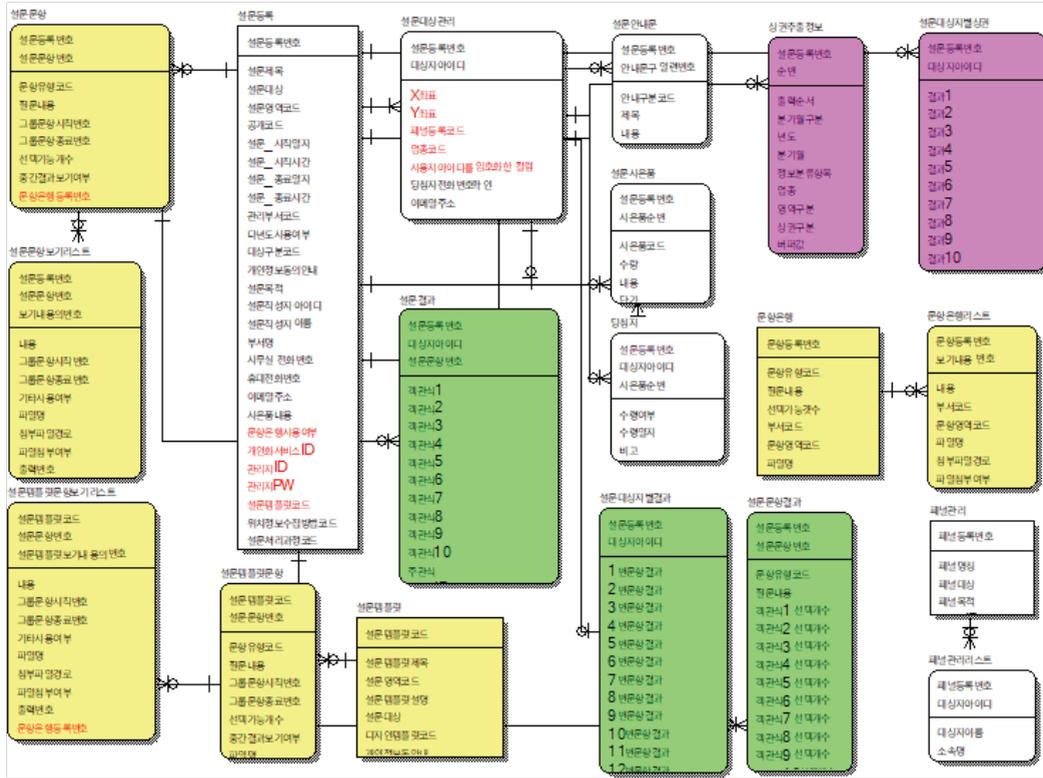
분석된 상권정보와 설문결과 연계 분석을 위

해 설문대상자들에 대한 X좌표와 Y좌표의 정보와 업종정보를 관리한다. 설문 대상자의 위치정보를 미리 저장하거나, 설문 시 설문자의 위치정보를 자동으로 입력받거나, 설문대상자의 주소검색을 통해 위치 정보를 입력 받을 수 있다. 그리고, 상권정보가 업종별로 구성되어 있어 업종에 대한 정보도 대상자 개인별로 관리하도록 한다. 사업자 번호가 있을 경우, 사업자 정보와 매핑이 가능하기도 하지만, 설문 대상자의 업종을 대상자별로 관리하도록 한다.

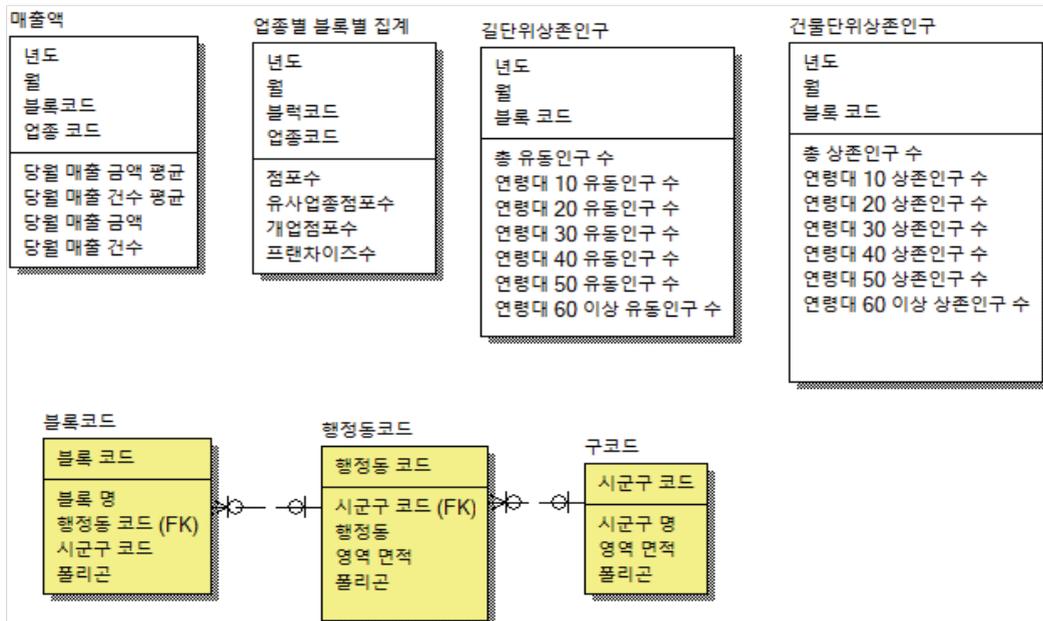
그림 3은 상권 서비스의 최소단위인 업종, 블록, 월을 기준으로 매출액, 업종, 길단위상존인구, 건물단위상존인구의 DB스키마이다. 그리고, 블록코드, 행정동코드, 구코드는 폴리곤으로 되어 있고, 영역의 포함관계를 외래키로 관리하고 있다. 집계 연산을 최소화하기 위해 행정동별, 시군구별 집계정보 테이블도 별도로 구성되어 있다.

3.3 설문정보와 상권정보 연계 방법

본 논문에서는 설문결과를 상권정보를 연계 분석하기 위한 설문대상자 개인에 대한 업종정보와 X, Y좌표 정보를 이용하여 상권정보와 매핑을 한다. 그림 4는 개인별 설문결과와 상권정보를 연계방법을 도식화한 것이다. 설문 대상자의 개인별 업종 정보는 상권정보와 1:1로 매핑이 되

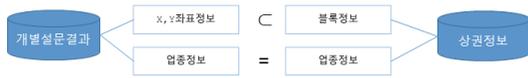


〈그림 2〉 설문조사 ERD(Entity Relation Diagram)



〈그림 3〉 상권정보 ERD(Entity Relation Diagram)

고, 개인별 X, Y좌표 정보는 블록코드 테이블의 폴리곤 정보를 이용하여 어떤 블록에 속하는지를 찾아내어 상권정보의 블록코드와 매핑한다. 매핑된 정보는 분석을 위한 2차원 테이블 형태로 설문결과리스트와 상권정보리스트로 표시할 수 있다.



〈그림 4〉 설문결과와 상권 연계 방법

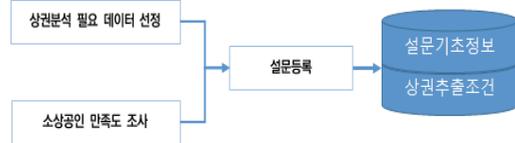
본 논문에서 개인별 설문결과와 매핑되어 추출되는 상권정보는 공간연산을 사용하여 많은 부하를 발생시키기 때문에 상권정보 추출 개수를 10개 제한하여 구성하였다.

3.4 모듈별 기능

본 절에서는 빅 데이터 기반의 상권정보와 연계한 설문시스템의 모듈별 기능에 대해서 설명한다. 제안하는 설문시스템은 상권서비스와 연계를 해야 하고, 사용자 편의성 향상을 위한 다양한 기능이 필요하다.

3.4.1 설문등록

일반적으로 상권분석에 추가적인 데이터가 필요하거나, 데이터 검증 및 소상공인의 의견이 필요할 때 설문을 실시하게 된다. 설문 결정 후 가장 먼저 시스템에 등록하는 것이 설문기본정보이다. 설문기본정보에는 설문의 제목, 목적, 대상, 개인정보 동의 내역 등을 등록한다. 그리고 설문개개인에 대한 위치 정보를 기반으로 상권정보와 연계분석 가능하도록 상권데이터 추출 조건을 입력한다. 상권데이터 추출 조건에는 기준년도, 선택기준, 세부기준, 정보 분류, 세부정보 분류, 업종, 세부업종분류, 영역구분이 있다. 이 조건은 결과 생성 시 개인별 위치정보를 기반으로 관련된 상권정보 추출 기준이 된다.



〈그림 5〉 설문등록 프로세스

3.4.2 문항관리

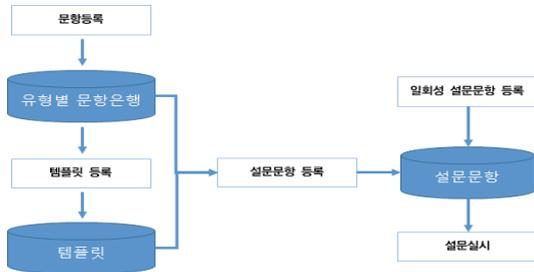
설문문항은 유형별 문항은행을 만들어 설문 의 일관성과 전문성을 유지하도록 한다. 문항은행에 등록된 질의문항을 목적에 맞게 조합하여 템플릿을 만들고, 이 템플릿을 선택하여 설문문항을 등록한다. 설문문항의 템플릿 사용을 통해 설문작성의 업무 효율을 높이고 일관성 있는 설문을 작성할 수 있다. 또한 동일한 템플릿으로 지속적으로 설문을 실시한 데이터는 데이터 분석 시 시계열 분석이 가능하다.

설문문항 작성 시 유형별 문항은행에 있는 문항을 사용하지만 설문 설계자의 관점에 따라 별도로 사용할 수 있게 하기 위하여 문항은행에 있는 문항을 조합하여 설문 문항을 등록할 수 있다. 이와 같은 설문문항은 문항은행에 등록된 문항은행 키를 참조하고 있어, 문항 은행에 있는 문항이 어떤 설문에서 사용되고 있는지를 분석할 수 있고, 설문별 특징을 분석할 수 있는 기초자료가 된다.

일회성 설문문항은 설문은행에 있는 설문문항을 참조하지 않고 일회성으로 설문을 만들어서 사용한다. 일회성 설문은 특수한 상황에서 발생하는 일에 대해서 설문을 하는 경우로 예를 들어, 코로나 발생에 따른 소상공인 임대료 조사에 등에서 사용할 수 있다.

데이터 검증 및 세분화된 집계 서비스 제공을 위해 문항 작성을 할 경우, 상권데이터에 있는 데이터를 설문문항으로 작성하여 설문을 실시한다. 설문이 완료된 후 설문에 대한 데이터의 집계와 상권에 있는 데이터를 비교하여 데이터의 정합성을 검증하고, 설문을 상권데이터 보다 작은 영역에서 집계를 하고, 집계된 비율로 상권데이터를

작은 영역으로 나누면 작은 영역에의 집계데이터를 추정할 수 있다.



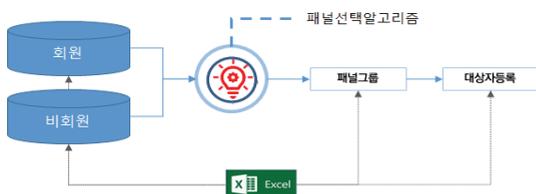
〈그림 6〉 문항등록 프로세스

3.4.3 대상자와 패널 관리

대상자는 회원과 비회원으로 구분되어 관리한다. 회원은 정책담당자 또는 전문가가 신용보증재단에 서비스 사용을 신청하여 허가 받은 사람들로 시스템의 전체 메뉴를 사용할 수 있는 권한을 갖는다. 비회원은 소상공인들로 주로 설문 대상자가 된다. 그러나 소상공인들이 개인화 서비스를 신청하게 되면 설문영역에서 개인화된 설문을 작성하고 운영할 수 있다.

회원과 비회원을 대상으로 소상공인을 대표하는 패널 선택 알고리즘을 사용하여 패널을 추출한 후 패널을 중심으로 설문조사가 가능하도록 한다. 패널로 등록된 대상자는 일정한 기간을 기준으로 지속적으로 설문을 실시해야 한다. 패널 대상자도 변동이 발생하지만 소상공인의 대표성을 잃지 않기 위해 삭제 시 기존의 패널 선택 알고리즘을 사용하여 삭제된 인원만큼 추가한다.

대상자들에 대한 관리의 편리성을 확보하기 위해 대상자들과 관련된 업무는 모두 엑셀 다운



〈그림 7〉 대상자관리 프로세스

로드 및 업로드가 처리되도록 한다. 비회원 입력 시 한명씩 입력하는 것보다 엑셀로 일괄 입력하는 것이 효율적이기 때문이다.

3.4.4 설문시행

설문을 시행하는 방법은 로그인을 기준으로 또는 회원여부를 기준으로 여러 가지 방법이 있다. 첫째, 가장 기본적으로 회원인 경우 로그인을 하면 자신과 관련된 설문조사 리스트가 조회되고 설문을 선택하여 설문을 진행하는 방법이다.

둘째, 설문관리자가 로그인 후 설문 대상자를 만나 인터뷰하며 설문을 진행하는 방법이다. 설문관리자는 대상자 한명 한명에 대해서 설문을 선택하고 질의 결과를 입력할 수 있는 권한을 갖고 있다.

셋째, 비로그인 방법으로 로그인 하지 않고 회원과 등록된 비회원에게 설문을 할 수 있는 방법으로 E-mail, SNS, SMS를 통해 발송된 URL로 접속하는 방법이다. 비로그인 설문대상자들의 설문 번호와 회원번호를 사용하여 인터넷정보원에서 제공하는 seed 암호화 알고리즘을 사용하여 단방향으로 암호화된 키 값을 만든다. 암호화된 키 값과 설문 URL값을 E-mail, SNS, SMS를 통해 대상자에게 전송한다. 전송된 메시지의 URL을 클릭하면 설문주소로 연결되어 설문을 실시하는 방법이다. 설문 서버에서는 설문주소에 있는 암호화된 키 값을 GET 방식으로 받아 암호화 키 값과 일치하는 대상자에서 찾아 설문 기간, 완료여부 등을 체크한 후 조건을 만족할 때 설문을 진행할 수 있다.

넷째, 등록되지 않은 비회원을 대상으로 게시판에 URL을 공유를 통해 설문을 진행하는 방법이다. URL을 통해 진행될 때는 접속IP 및 위치정보를 저장하여 중복 설문에 대해서 점검할 수 있도록 한다. URL을 접속하여 설문을 진행하는 사람들은 대상자 관리를 위해 임의의 임시번호가 부여된다.

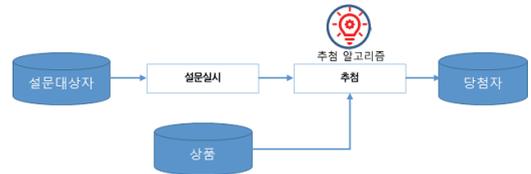
다섯째, 오프라인으로 종이설문을 하는 방법이다. 종이설문을 완료하면, 설문기본 정보와 문학을 등록하고, 엑셀에 오프라인 설문결과 값을 입력한 후 엑셀 업로드 기능을 통해 시스템에 입력할 수 있다.

여섯째, 중간관리자가 관리하는 방법으로 설문 등록시 등록된 관리자 아이디와 패스워드가 암호화된 키 값으로 URL과 함께 제공되어, 이 주소를 클릭하여 설문을 실행하는 방법이다. 중간관리자가 인터뷰하며 설문을 진행한다. 두 번째 방법과 다른 것은 두 번째 방법은 관리자가 로그인하여 설문을 실행하는 방법이라면, 이 방법은 중간관리자는 설문이 실행되는 URL과 관리자 아이디 값과 비밀번호 값이 암호화된 값을 사용하여 시스템에 접속하여 설문을 실행하는 방법이다. 관리자는 중간관리자가 설문을 실행할 수 있는 값을 수시로 변경해 가면 중간관리자가 시스템에 접속하는 것을 관리할 수 있다.

3.4.5 사은품 관리

사은품은 설문을 등록 시 사은품에 대해서 등록을 하고, 설문을 실시한 대상자들에 한해서 사은품 추천을 한다. 사은품 추천은 추천 알고리즘

을 통해 랜덤하게 추천되며, 중복 추천이 되지 않도록 구성이 된다. 당첨자들은 회원/비회원 정보에 등록이 되어 있는 연락처 및 이메일 주소를 사용하여 당첨자에게 당첨 사실을 통보한다.



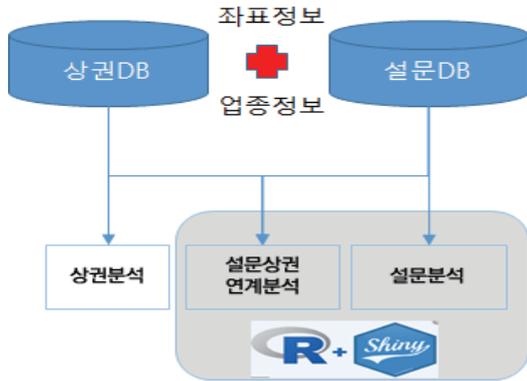
〈그림 9〉 사은품관리 프로세스

3.4.6 빅 데이터 기반의 상관정보와 설문 연계 분석

본 설문시스템에서는 설문 문항 간 분석과 상관정보와 설문문항 간 연계한 분석이 가능하다. 설문 문항 간 분석은 T-test, 상관분석, 회귀분석 등을 R-shiny를 통해서 수행하고, 상관정보와 설문문항 간 분석은 설문 대상자의 위치정보를 기준으로 상관정보와 매핑하여 상관과 설문데이터를 1:1 형태로 만들고, 분석기법을 선택한 후 R-shiny로 데이터를 전송하여 분석한다[14]. 즉, 설문시스템에서 선택된 데이터를 R-shiny 서버로 전송하고, R-shiny는 get방식의 json형식으로 데이터를 받아 R의 분석 모듈을 사용하여 분석한다.

로그인	설문프로세스	설문대상
로그인	설문시스템 로그인 → 설문실시	회원
	비회원선택 → 인터뷰 설문	등비회원
비로그인	암호화 → Email, SNS, SMS → 대상자 → 설문실시	회원 등비회원
	암호화 → 게시판 → 불특정 대상자	미비회원
오프라인	종이설문실시 → 취합 → 엑셀업로드	미비회원
	URL → 인터뷰 설문	

〈그림 8〉 설문실행 프로세스



〈그림 10〉 상권 및 설문 분석 프로세스

IV. 빅 데이터 기반 상권서비스 연계 분석 설문조사시스템 구현

본 절에서는 시스템 구현환경과 상권정보 연계를 위해 구현된 설문조사 시스템의 기능을 중심으로 설명한다.

4.1 구현환경

본 논문에서는 전자정부프레임워크 3.6 기반으로 DB는 Oracle 12g로 개발하고, 분석도구로 R-shiny 3.8을 사용하여 구축한다. WAS로 WebToB 4.0과 Jeus 4.0을 사용한다. 그리고 지도서비스를 위해 Kakao Map Api를 사용하고[16], 위치정보 추출을 위해 HTML Geolocation API를 사용한다 [15].

4.2 연계 분석할 상권정보 추출 조건 선택

설문등록 시 설문과 연계 분석할 상권정보 추출 조건을 선택하고, 설문을 완료하면 개인별 위치정보를 기준으로 선택된 상권정보를 추출한다. 이를 위해, 먼저 연계분석 할 상권 정보의 기준 연도를 선택한다. 선택한 연도를 기준으로 추출한 정보의 분기별 또는 월별 기준을 선택한다. 선택한 기준에 따라 월과 분기를 입력한다. 그리고

정보 분류인 개폐업수(률)을 선택한 후 세부 정보 분류인 개업수를 선택하면 정량적인 정보 선택 기준을 마친다. 그런 후 업종을 선택하기 위해 오른쪽에 있는 업종 콤보 박스를 선택하고, 선택한 업종에 대한 분석하고 싶은 세부업종을 선택한다. 마지막으로 분석 범위를 선택하기 위해 영역을 블록, 행정동, 구, 위치 중 하나를 선택한다. 블록, 행정동, 구는 개인별 위치정보가 포함되어 있는 블록, 행정동, 구를 사용하지만, 위치는 영역이 고정되어 있지 않다. 즉 설문대상자의 위치 정보를 기준으로 원 형태로 몇 m 떨어진 영역까지 포함하여 통계 계산을 해야 하는지를 버퍼에 입력한다. 그림 11은 상권정보 추출조건 입력화면으로 아래에 있는 상권정보는 설문과 연계를 위해 선택된 항목이다.

상권데이터 추출 설정

기준년도: 2020년 | 업종: 의식업

기준선택: 분기 | 세부업종: 한식음식점

세부 기준 선택: 1분기 | 영역구분: 위치

정보분류: 개폐업수(률) | 버퍼: 30

세부 정보 분류: 개업수

저장

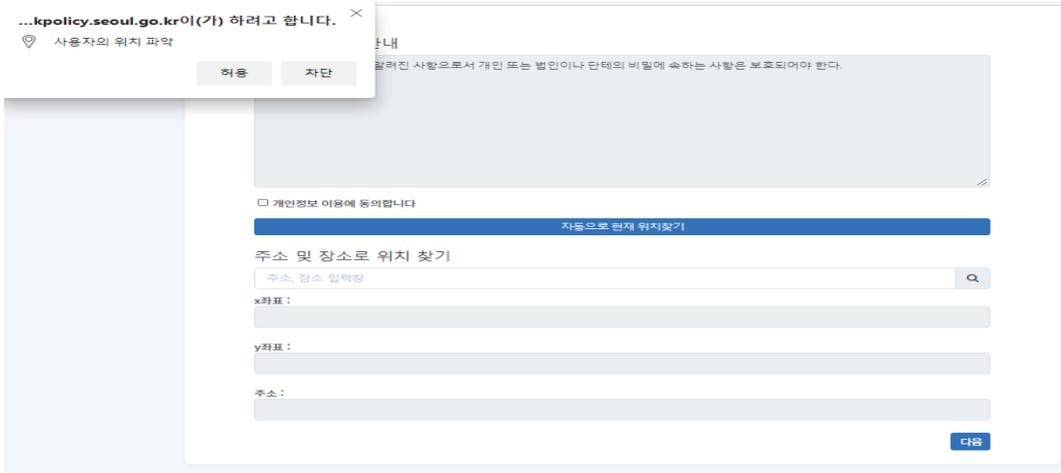
*상권정보는 최대 9개 까지 연계 가능합니다.

순서	기준	년	분기	정보분류	업종	영역구분	상권구분	버퍼	삭제
1	분기	2019	2	전체점포수	대시인서비스업종 전체	블록			X

〈그림 11〉 상권정보 추출 조건 입력 화면

4.3 설문조사 진행 중 위치정보 수집

위치기반 설문조사를 위해 설문대상자 입력 시 사업장의 좌표정보를 입력받거나, 설문 실행 중 설문대상자 사업장의 위치 정보를 입력받는다. 이를 위해 설문 실행 시 설문자의 위치 정보를 자동으로 끌어 오거나, 건물 또는 주소를 조회하여 위치 정보를 선택한다. 설문자의 위치 정보를 자동으로 끌어 오기 위해 html5에서 제공하는

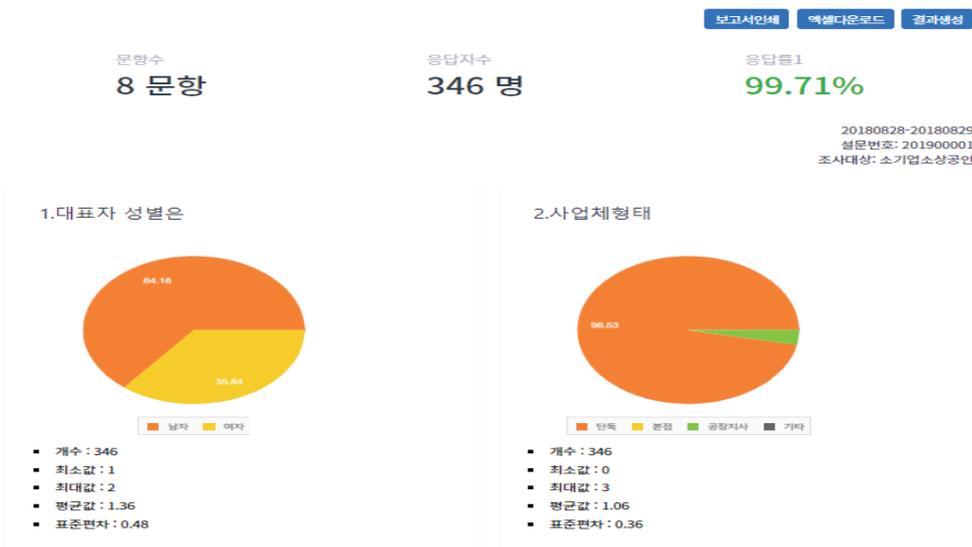


〈그림 12〉 설문 진행 중 위치정보 수집 화면

geolocation 함수를 사용하고, 위치 정보를 건물명 또는 주소로 조회하여 입력 받기 위해 kakao에서 제공하는 API를 사용한다. 그림 12는 설문 진행 중 위치정보를 수집하는 화면으로 자동으로 위치정보를 수집하게 설정을 하면 설문 접근시 위치정보 사용을 여부를 지정하는 팝업창이 뜬다. 위치정보 허용을 하고 자동으로 현재 위치 찾기 버튼을 클릭하면 주소정보와 좌표정보가 HTML Geolocation API를 사용하여 저장된다.

4.4 설문 결과

설문결과는 설문이 완료되면 그림 13의 오른쪽 상단에 있는 결과생성 버튼을 클릭하여 생성한다. 설문결과 생성으로 설문문항을 기준으로 한 결과와 설문개개인에 대한 결과가 생성된다. 설문문항별 결과는 그림 13과 같이 설문 전체의 요약을 한 후 개별 문항에 대한 결과를 그래프 형태로 제공한다. 그리고 설문개개인에 대한 결



〈그림 13〉 설문문항별 결과

과는 설문 개인별 결과와 상권정보 추출 기준으로 추출한 데이터를 엑셀로 다운받는 형태로 제공한다. 그림 14는 개인별 설문결과와 연계된 상

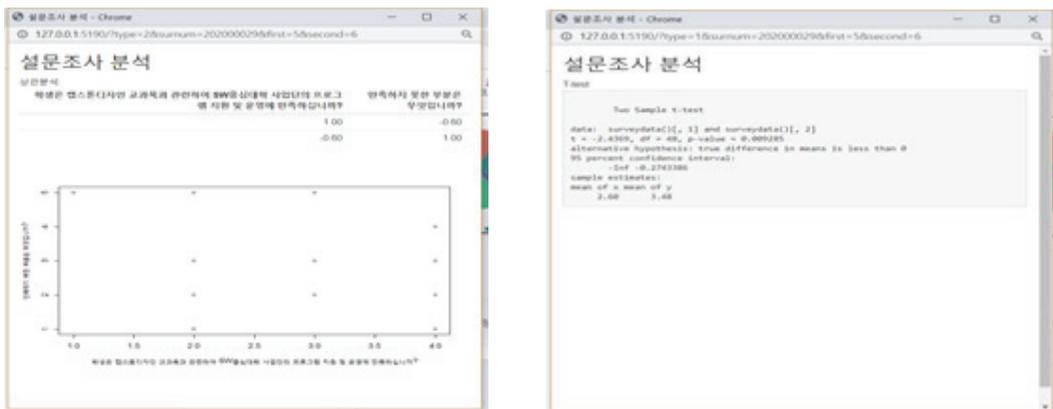
권정보의 점포수, 주거인구, 영업기간의 정보를 1:1로 매핑된 형태의 엑셀로 다운로드 받은 것이다. 이와 같이 매핑된 설.

상호	X좌표	Y좌표	대표자 성별은	사업제형태	사업장형태	임차료월세	인건비(만원)	외부종사자수	창업자금	프랜차이즈여부	제정표수대시	19년 2분기주거	영업기간
몽방노래연습장	212179.2090	448796.8160	2	1	3	200	0	0	20000	2	24	298	2.3
헤이존	212715.9748	449069.7792	2	1	3	160	250	1	10000	2	32	298	2.3
크린토피아	212932.8392	449864.2520	2	3	3	77	0	0	7000	1	22	353	2.7
대동공신	211612.1343	448579.7146	1	1	3	70	4000	27	5000	2	72	491	2.3
소우영크린티다	211278.4720	448490.9055	1	1	3	110	0	0	10000	2	10	400	2.4
원조판 전문점	211227.3048	448894.7960	1	3	3	500	400	1	30000	2	24	192	2.5
송과컴퓨터회계	209946.9573	444534.2426	1	1	3	495	1000	3	10000	2	31	331	2.5
나라사랑집류터	210914.5653	443362.9103	1	1	3	100	1000	4	4000	2	59	360	2.4
현대자동차공회	210964.6307	443919.3998	1	1	3	1210	1000	3	20000	2	13	362	2.4
대신자동차정비	210812.8329	444434.8008	1	1	3	1500	4000	11	20000	2	13	276	2.4
사탄헤어살롱	210277.4510	443524.5926	2	1	3	280	150	1	10000	1	93	124	2.1
문신기록	210554.3114	443760.1571	1	1	3	10	500	3	10000	2	260	276	2.4
영채카	210804.9120	444027.5332	1	1	3	500	400	1	10000	2	40	276	2.4
화산실업	210338.9170	447771.6233	1	1	3	200	500	3	5000	2	33	123	2.6
정찬컴퓨터교육	207420.6149	445350.1515	1	1	3	200	500	2	5000	2	12	254	2.4
동구연아샵	210515.3128	443756.9488	1	1	3	50	0	0	5000	2	260	276	2.4
미주헤어샵	213608.8968	449904.8635	2	1	3	150	0	0	5000	2	16	197	2.7
세븐데이즈 뷰티	210566.9441	447971.9155	2	1	1	0	800	4	10000	2	19	123	2.6
미디어릭스	210589.0305	442308.7586	1	1	3	150	0	0	5000	2	153	124	2.1
원대DVD방	210901.2201	448986.9593	1	1	3	100	0	0	10000	2	46	192	2.5
리코벌처스	211307.0026	443121.8968	1	1	3	100	0	0	3000	2	10	360	2.4
환산컨설팅	210589.0305	442308.7586	1	1	3	60	0	0	5000	2	153	124	2.1
스칼라	212901.8548	444400.8905	2	1	3	80	0	0	3000	2	16	245	2.4
경진종합설비	213624.0390	444044.5712	1	1	3	80	0	0	5000	2	4	369	2.1
신미디어	209107.8044	444499.7726	1	1	3	60	0	0	5000	2	33	361	2.3
문화회백는날	211943.2786	443871.5769	2	1	3	60	0	0	3000	2	25	362	2.4
아름소리음악원	215115.3465	450122.9608	2	1	3	100	0	0	10000	2	32	62	2.6
에이클레식	213464.8320	451222.8310	2	1	3	100	0	0	10000	2	16	142	2.3

〈그림 14〉 개인별 설문결과와 좌표기준으로 연결된 상권정보



〈그림 15〉 분석 방법 및 분석 항목 선택 화면



〈그림 16〉 상관분석과 T-test 분석 결과



〈그림 17〉 회귀분석 결과

4.5 설문 간 분석 및 빅 데이터 기반 상관정보 연계 분석

설문결과와 상관정보는 2차원 형태의 테이블 형태로 구성되어 다양한 데이터 분석 방법을 적용할 수 있다. 본 논문에서는 T-test 분석, 상관분석, 회귀분석이 가능하도록 시스템을 구축하였다. 분석을 위해 먼저 분석 방법을 선택한 후 문항과 문항 또는 문항과 상관정보를 대상으로 분석하고자 하는 항목을 선택한다. 그림 15은 분석하고자 하는 분석 방법 및 항목 선택하는 화면이다. 그림 16는 두 항목간의 연관성을 알아보는 상관분석과 차이를 분석하는 T-test 분석 결과를 R-shiny를 통해 출력된 화면이고, 그림 17은 미래를 예측할 수 있는 회귀분석 결과를 출력한 화면이다.

V. 결 론

본 논문에서는 빅 데이터 기반 상관분석 서비스의 한계를 극복하기 위해 상관정보와 연계 분석이 가능하고 사용자 편의성 향상을 위해 다양

한 형태의 설문이 가능한 위치기반 설문조사 시스템을 설계 및 구현하였다.

상관정보와 연계분석을 위해 첫째, 설문등록 시 연계분석이 가능한 상관정보를 선택하도록 설계하였다. 둘째, 설문대상자의 위치정보를 수집하기 위해 대상자 입력 시 또는 설문 시 위치정보 수집이 가능하도록 하였다. 셋째, 설문대상자의 위치정보와 상관의 영역정보를 사용하여 1:1로 매핑한 설문결과와 상관정보를 다양한 분석 가능하도록 R-Shiny를 연계한 시스템을 구축하였다. 구축된 시스템은 상관정보와 연계 분석이 가능하여 상관데이터의 정합성 검증, 데이터 분석의 한계, 집계 데이터의 세분화가 가능하게 되었다.

또한, 다양한 형태의 설문 기능을 추가하여 첫째, 권한관리를 계층적으로 세분화 하였다. 둘째, 설문 유형별 문항은행을 만들어 설문의 일관성과 전문성을 유지하도록 하였다. 셋째, 패널 그룹을 지정하여 관리할 수 있도록 시스템을 구성하여 설문의 신뢰성을 높였고, 시계열적인 설문 분석이 가능하도록 하였다. 마지막으로, 설문의 편의성을 높이기 위해 설문 템플릿을 구성하여 편리하게 설문을 구성할 수 있도록 하였다.

본 논문에서 제안한 설문조사시스템은 빅 데이터 상권정보 분석 서비스의 한계점을 보완하고, 분석 서비스를 확장할 수 있는 기반을 마련하였다. 향후 연구과제로는 설문과 상권서비스에 대한 다양한 연계 방법을 찾아보고, 연계된 정보에 대한 빅 데이터 분석 방법을 수행하는 것이다.

참 고 문 헌

[1] N. I. Sung, J. K. Kim, "Entry and Exit in Korea's Service Industries and Their Determinants: An Empirical Analysis," *The Korean Journal of Industrial Organization*, vol. 28, pp. 37-88, 2020.

[2] 강만수, 윤재웅, 정경훈, 위평량, "상권시스템으로 본 서울 영세자영업 통계 및 젠트리피케이션 양상 분석과 시사점", 서울신용보증재단 소상공인 정책연구센터, 연구보고서, 2019.12.

[3] Seoul Metro City, <https://golmok.seoul.go.kr/main.do>.

[4] J. H. Kong, B. C. Yun, et al., "Commercial district selection through big data analysis: Restaurant business", *Proc. of Korea Software Congress*, pp. 1938-1940, Dec. 2017.

[5] H. M. Kang, S. K. Lee, "Analyzing Growth Factors of Alley Markets Using Time-Series Clustering and Logistic Regression," *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, vol. 37, pp. 535-543, 2019.

[6] H. C. Kim, Y. S. An, "A Monitoring the Dynamic Change of Seoul's Side Street Trade Areas Using Density and Diversity of Stores," *Seoul Studies*, vol. 20, pp. 149-170, 2019.

[7] H. G. Kim, "Developing a Big Data Analysis Platform for Small and Medium-Sized Enterprises," *Journal of the Korea Society of Computer and Information*, vol. 25, pp. 65-72, 2020.

[8] Micro Enterprise And Market Service, <https://sg.sbiz.or.kr/> .

[9] Gyeonggi Market Revitalization Agency, <https://sbiz.gmr.or.kr/main.do> .

[10] <https://www.qualtrics.com/experience-management/research/> .

[11] <https://docs.google.com/forms/u/0/> .

[12] <https://office.naver.com/> .

[13] <https://ko.surveymonkey.com/> .

[14] <https://rstudio.com/> .

[15] <https://developers.google.com/maps/documentation/javascript/examples/map-geolocation>.

[16] <https://apis.map.kakao.com/> .

저 자 소 개



이 원 철(Won-Cheol Lee)

- 2000 강원대학교 행정학과 (학사)
- 2002 강원대학교 컴퓨터과학과 (석사)
- 2007 강원대학교 컴퓨터과학과 (박사수료)
- 2019~현재 한림대학교 소프트웨어중심대학사업단 산학협력중점교수
- 관심분야 : 상권데이터 분석, 데이터베이스 설계 및 튜닝, 접근제어 설계, 업무시스템 개발, 머신러닝 등

E-mail: woncheol@kangwon.ac.kr



강 만 수(Man-Su Kang)

- 1997 강원대학교 경영학과 (학사)
- 2000 강원대학교 경영학과 (석사)
- 2011 강원대학교 경영학과 (박사)
- 2013~2016 신용보증재단중앙회 연구원
- 2016~현재 서울신용보증재단 소상공인 정책연구센터 데이터인프라팀장
- 관심분야 : 금융 빅 데이터, 상권분석서비스, 소상공인 컨설팅시스템, 다차원 정보수집 시스템 개발 등

E-mail: jazz@kangwon.ac.kr



김 진 호(Jinho Kim)

- 1982 경북대학교 전자공학과 (학사)
- 1985 KAIST 전산학과 (석사)
- 1990 KAIST 전산학과 (박사)
- 1990~현재 강원대학교 컴퓨터과학과 교수
- 2020~현재 강원대학교 빅 데이터·융합보안사업단 단장
- 관심분야 : 대용량 빅 데이터 저장 및 처리, 하둡/맵리듀스 분산/병렬처리 기술, 빅 데이터 분석 기법, 데이터 마이닝, 클라우드 컴퓨팅, 데이터 웨어하우스, OLAP 다차원 분석, 데이터베이스 시스템 개발

E-mail: jhkim@kangwon.ac.kr