

공간 빅데이터의 개념 및 요구사항을 반영한 서비스 제공 방안 Providing Service Model Based on Concept and Requirements of Spatial Big Data

김근한* · 전철민** · 정휘철*** · 윤정호****

Kim, Geun Han · Yoon, Jeong Ho · Jun, Chul Min · Jung, Hui Cheul

要 旨

본 연구에서는 빅데이터와 공간 빅데이터 선행연구들을 기반으로 공간 빅데이터를 빅데이터를 구성하는 하나의 구성요소로 인식하고, 위치정보를 이용하여 공간화 할 수 있으며, 시계열 변화에 따라 계속적으로 누적되는 모든 데이터들과 이를 이용할 수 있는 활용체계를 공간 빅데이터라 정의하였다. 따라서 공간 빅데이터는 기존 빅데이터와 분리하여 구분할 것이 아니라, 기존 빅데이터를 구성하는 하나의 구성요소로서 이해하고, 이러한 활용체계 안에서 공간 빅데이터의 활용방안을 검토해야 한다. 본 연구에서는 공간 빅데이터가 제공해야 하는 서비스 요구사항들을 제시하였다. 공간정보를 포함한 공간 빅데이터는 기본적으로 다양한 공간분석이 가능해야 하고, 기존에 구축된 공간정보와 향후 구축될 공간정보까지 고려할 수 있는 서비스 고려가 필요하다. 시간의 흐름에 따른 위치별 시계열 변화의 탐지는 물론 공간정보의 속성정보들을 이용하여 다양한 빅데이터 관련 분석이 가능해야 한다. 공간정보가 아닌 빅데이터 또한 공간정보와 연계하여 공간 분석이 가능해야 한다. 이러한 공간 빅데이터 요구사항들을 만족시키기 위해 다양한 형태의 빅데이터들과 공간 빅데이터의 연계가 가능한 분석 서비스 제공을 위한 샘플링 포인트 생성 및 속성정보 추출 방안을 제시하였다. 이러한 빅데이터와 연계된 공간정보의 활용 증대는 공간정보 산업 및 기술발전에 크게 기여할 수 있을 것이라 판단된다.

핵심용어 : 빅데이터, 공간 빅데이터, 공간 빅데이터 서비스, 공간 빅데이터 요구사항, 샘플링 포인트 속성정보 추출

Abstract

By reviewing preceding studies of big data and spatial big data, spatial big data was defined as one part of big data, which spatialize location information and systematize time series data. Spatial big data, as one part of big data, should not be separated with big data and application methods within the system is to be examined. Therefore in this study, services that spatial big data is required to provide were suggested. Spatial big data must be available of various spatial analysis and is in need of services that considers present and future spatial information. Not only should spatial big data be able to detect time series changes in location, but also analyze various type of big data using attribute information of spatial data. To successfully provide the requirements of spatial big data and link various type of big data with spatial big data, methods of forming sample points and extracting attribute information were proposed in this study. The increasing application of spatial information related to big data is expected to attribute to the development of spatial data industry and technological advancement.

Keywords : Big Data, Spatial Big Data, Spatial Big Data Service, Spatial Big Data Requirements, Extraction of Attribute in Formation of Sampling Points

Received: 2016.11.30, revised: 2016.12.09, accepted: 2016.12.16

* 정희원 · 서울시립대학교 공간정보공학과 박사과정 & 한국환경정책평가연구원 연구원(Member, Ph. D. Student, Department of Geoinformatics, University of Seoul & Researcher, Korea Environment Institute, ghkim@kei.re.kr)

** 서울시립대학교 공간정보공학과 교수(Professor, Department of Geoinformatics, University of Seoul, cmjun@uos.ac.kr)

*** 한국환경정책평가연구원 연구위원(Researcher, Korea Environment Institute, hchjung@kei.re.kr)

**** 교신저자 · 한국환경정책평가연구원 연구위원(Corresponding Author, Researcher, Korea Environment Institute, jhyoon@kei.re.kr)

1. 서 론

1.1 연구배경 및 목적

소셜미디어 및 인터넷 웹 서비스의 발전과 스마트기 및 IoT 기반 센서기술의 발전으로 다양한 종류의 데이터들이 생성되고 있다. 기존 분석 및 활용에 고려되지 않았던 이러한 데이터들을 분석 및 활용하여 새로운 가치를 창출하는 빅데이터가 다양한 산업 및 연구 분야에서 큰 이슈로 부각되면서 다양한 분야에서의 활용이 급증하고 있다. 빅데이터가 새로운 가치를 창출 할 수 있는 중요한 자원으로 인식되고, 이슈화 되면서 공간정보 분야 또한 공간 빅데이터의 개념과 서비스 및 활용과 관련된 연구들이 지속적으로 연구되고 있다.

하지만 아직까지 공간 빅데이터의 개념 정립이 명확하지 않고, 공간 빅데이터를 이용한 분석 방법 및 서비스 방안에 대한 연구는 아직까지 미흡하다. Kim et al.(2013)은 공간 빅데이터 기반의 사회적 약자관리 침수재해 모핑터링에서 사회적 약자의 위치정보와 기상정보를 이용하여 침수예상주택들을 선정하였으며, 도로 침수에 따른 우회경로 설정시스템 구축을 위해서는 상습침수 지역과 도로의 공간정보를 이용하여 우회경로를 선정하는 분석을 수행하였다. Kim et al.(2014)은 건물에너지 빅데이터 분석을 위해 130만건의 개별 건축물별 에너지 사용량 데이터를 KLIS(한국토지종합정보망)에서 추출한 개별 필지 데이터와 통합하여 분석에 이용하였다. 이러한 데이터를 이용하여 각종 통계분석 및 핫스팟 분석과 같은 공간 분석을 수행하였다. 이렇게 몇몇 연구에서 공간 빅데이터 분석과 관련된 예시들을 제시하고 있지만, 이러한 공간 빅데이터 분석들은 기존의 공간분석과 별반 다르지 않다.

이에 본 연구에서는 빅데이터와 공간 빅데이터의 연구 사례들을 검토하고 이를 기반으로 공간 빅데이터의 정의를 내리고, 공간 빅데이터가 제공해야하는 요구사항들을 제시하였다. 그리고 이러한 요구사항들을 만족할 수 있는 공간 빅데이터 서비스 구현 방안을 제시하였다.

1.2 연구방법

우선 빅데이터 선행연구들에서 제시한 빅데이터에 대한 정의들을 살펴보고, 이러한 빅데이터 정의들의 그룹화를 통해 빅데이터 정의에서의 공간 빅데이터의 위치와 역할을 살펴보았다. 그리고 국내외의 공간 빅데이터 관련 연구 사례들을 분석하여 기존 공간 빅데이터의 개념 및 특징을 살펴보았다. 이러한 빅데이터와 공간 빅데이터들의 선행연구를 기반으로 공간 빅데이터의 정

의를 내렸으며, 공간 빅데이터가 제공해야 하는 서비스 요구사항들을 제시하였다. 그리고 이러한 요구사항들을 만족시킬 수 있는 공간 빅데이터의 서비스 구현 방안을 빅데이터와의 연계를 고려하여 제시하였다.

2. 공간 빅데이터 개념 정립

2.1 빅데이터의 개념과 특징

빅데이터의 개념은 현재까지도 Table 1과 같이 연구자, 연구기관마다 다양하게 해석하고 있다. 우선 기존의 저장·관리·분석의 체계로는 감당할 수 없는 양의 데이터 중심의 빅데이터 정의를 살펴보면 McKinsey & Company는 빅데이터를 수집, 저장, 관리, 분석을 수행하는 전형적인 데이터베이스 소프트웨어의 능력을 초과하는 규모의 데이터로 정의하였다(James and Michael, 2011). Gartner는 데이터의 양이 많고, 데이터 형태가 다양하며, 데이터 생성 속도가 빠른 데이터로 빅데이터를 정의하였다(Beyer and Laney, 2012). 국내에서는 Samsung Economic Research Institute에서 빅데이터를 기존의 관리 및 분석 체계로는 감당할 수 없을 정도의 거대한(100TB이상) 데이터의 집합으로 정의 내렸다(Park et al., 2013).

그리고 빅데이터를 통해 가치있는 데이터의 추출까지 고려한 빅데이터의 정의를 살펴보면 Oracle은 방대한 데이터(volume), 빠른 데이터 생성 속도(velocity), 다양한 종류의 데이터(variety), 데이터들로부터 얻는 가치(value)의 네 가지 특성을 만족하는 데이터를 빅데이터로 정의하였다(Dijcks, 2014). 여기서 가치란 빅데이터에서 데이터의 양만 고려할 것이 아니라 기존에 다루지 않았던 비전통적 데이터들을 분석하여 의사결정을 이끌어낼 수 있는 새로운 가치로 정의 내렸다. TeraData는 빅데이터를 기존의 전통적 데이터부터 반정형, 비정형의 복잡(complexity)하고 방대한 데이터 뿐만 아니라 빠른 생성 속도와 빠른 갱신 주기(frequent interval)를 가지고, 이러한 데이터들로부터 얻는 가치 있는 데이터를 빅데이터로 정의하고 데이터의 가치를 강조하고 있다(Teradata, 2013).

또한 방대한 양의 데이터를 수집, 관리, 분석, 활용하는 기술과 사람을 포함한 조직까지 아우르는 하나의 체계를 빅데이터로 정의하는 연구자, 연구기관도 있다. IDC는 다양한 종류의 대규모 데이터를 초고속 수집, 발견·발굴, 분석하여 가치 있는 정보를 경제적으로 추출하는 차세대 기술 및 아키텍처를 빅데이터 기술이라고 정의하였다(Gantz and Reinse, 2011). 그리고 Nomura Research Institute는 빅데이터를 처리할 수 있는 인제

Table 1. Various definitions of big data

Source	Definition
McKinsey&Company (James and Michael, 2011)	Dataset whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze
IDC (Gantz and Reinsel, 2011)	A new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling high-velocity capture, discovery, and/or analysis
Gartner (Beyer and Laney, 2012)	Data with vast amount of volume, variety in format, high velocity, and complexity
Samsung Economic Research Institute (Park et al., 2013).	Ruinous amount of data beyond the existing management and analysis system
Oracle (Dijcks, 2014)	Data satisfying a certain Volume, Velocity, Variety, Value
TeraData (2013)	Valuable semi-structured and unstructured data with high velocity in creation, massive volume, and short frequent interval derived from traditional data
Nomura Research Institute	Defines 3 element of big data: organization or the talented able to process big data, processing and aggregation of data, analysis techniques, data resource
National Council on Informatization Strategies (2011)	Extracting valuable data through analysis and application of massive data, and show active response or predict change based on the produced data
Kim (2014)	A massive amount of data with various formats and fast speed of creation requiring a new management and analysis method
Kim et al. (2013)	System capable to produce, collect, analyze, and applicate valuable data through integration of various data
Yi et al. (2014)	Ruinous amount of data beyond the existing management/analysis system and its techniques and tools
Yu et al. (2014)	Management, analysis, and application system including structured and unstructured data

·조직, 데이터 처리·축척·분석기술, 데이터 자원 등을 데이터와 데이터처리기술 및 이러한 데이터를 이용하는 인재를 빅데이터 3요소로 정의하였다(Kim et al., 2013). National Council on Informatization Strategies (2011)와 Kim et al.(2013), Kim(2014), Yi et al. (2014), Yu et al.(2014)는 빅데이터를 기존의 관리 및 분석 체계로는 감당할 수 없는 대용량 데이터를 활용·분석하여 가치 있는 정보를 추출하고, 생성된 지식을 바탕으로 능동적으로 대응하거나 변화를 예측하기 위한 정보화 기술 및 활용체계라고 정의하였다.

이러한 기존 선행 연구들의 빅데이터 정의와 특징들을 살펴보면 빅데이터는 기존의 환경 체계에서는 방대한 양의 데이터베이스의 정형 데이터뿐만 아니라 소셜 미디어, XML, HTML 같은 웹정보, 인터넷 게시물, 텍스트, 동영상, 이미지, 위치정보(공간정보)와 같은 다양한 비정형, 반정형 데이터들을 가리켰으나, 점차 빅데이터 범위가 확대되어 방대한 양의 데이터뿐만 아니라 데이터들로부터 추출하는 의미 있는 가치와 이러한 데이터를 다루는 인재 및 조직과 기술 및 활용체계까지 포함하여 빅데이터의 의미가 확대되고 있었다. 따라서 이러한 정의들을 살펴보면 공간 정보를 포함한 공간 빅데이터 또한 빅데이터의 일부로 판단하는 것이 타당할 것으로 판단되며, 공간 빅데이터는 다른 유형의 빅데이터들과의 연계할 수 있는 기술과 활용체계까지 고려해야 할 것으로 판단된다.

2.2 공간 빅데이터의 개념과 특징

빅데이터와 마찬가지로 공간 빅데이터 또한 명문화된 정의가 내려져 있지 않으며, Table 2와 같이 연구자마다 다양한 정의를 내리고 있다. 특히 국외 연구보다는 국내를 중심으로 2013년부터 공간 빅데이터와 관련된 연구들이 진행 중이었으며, 이러한 공간 빅데이터 관련 연구들에서 내린 정의들을 살펴보면 공간 빅데이터는 크게 세 가지로 구분할 수 있다.

Ahn et al.(2013)와 같이 데이터의 관리·처리·분석에 있어서 현재의 컴퓨팅 시스템 용량을 초과하는 공간데이터를 지칭하거나, Kim et al.(2013), Kim and Yoon(2013), Kim et al.(2014), Lee et al.(2014), Yu et al.(2014)와 같이 공간정보들과 공간정보와 연계·융합하여 공간데이터화 할 수 있는 빅데이터를 공간 빅데이터로 정의하고 있다. 그리고 Kim et al.(2014)는 공간 빅데이터를 공간적 속성과 함께 시간적 속성을 보유한 방대한 기록들의 집합이라 하였다. 이는 공간적인 위치정보마다 시계열 변화에 따른 정보의 변화까지도

Table 2. Various definitions of spatial big data

Source	Definition
Kim et al. (2013)	Spatial data fused with big data including administrative data from public sector, Twitter, internet information etc. or data able of spatial datafication
Kim and Yoon (2013)	Big data manufactured to a new format of information related with spatial information
Ahn et al. (2013)	Spatial data difficult to handle with the current system in the means of amount, processing speed, and variety
Kim et al. (2014)	Information of an action including temporal and spatial attributes. and a vast set of microscopic records
Kim (2014)	Spatial information and its techniques applied to big data analysis to extract a more fast, precise, and valuable analysis result
Lee et al. (2014)	Fusing various big data based on a place or location or big data with spatial information
Kim et al. (2014)	Spatial data and spatial informatization of big data
Yu et al. (2014)	Extraction of valuable information through effective collection, storage, and management of structured, unstructured, and antiformal spatial big data and integration /analysis of real time information

고려할 수 있어야 한다. 마지막으로 Kim(2014)은 공간 빅데이터를 다양한 공간정보들의 공간분석 기법을 빅데이터 분석과 접목할 수 있는 체계 또는 공간정보를 공간 빅데이터로 정의내리고 있었다.

이러한 선행연구에서 내린 공간 빅데이터의 정의와 관련하여 범위를 살펴보면 방대한 공간데이터 자체와 공간정보와 연계할 수 있는 빅데이터, 그리고 시계열 정보까지 포함한 공간데이터의 집합으로 정의내릴 수 있다.

이는 빅데이터의 정의 및 특징을 선행연구에서 살펴본 바와 같이 기존 관계형 데이터베이스와 같은 정형 데이터와 XML, HTML과 같은 반정형 데이터, 그리고 텍스트, 이미지, 동영상, 센서데이터, 공간정보(위치정보)와 같은 비정형 데이터로 구성되어 있다. 즉 Fig. 1과 같이 기존 빅데이터에 위치정보를 포함한 공간정보 뿐만 아니라 다양한 종류의 데이터들이 빅데이터 정의에 포함되어 있기 때문에 래스터, 벡터 포맷의 공간정보와 위성영상 등 공간정보 또한 빅데이터의 일부로 정

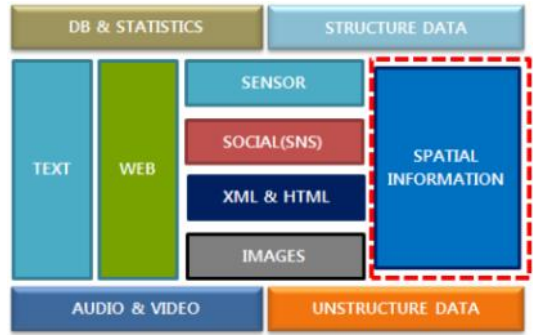


Figure 1. Various types of big data including spatial information

의 내리는 것이 타당하다고 판단된다. 그리고 이러한 공간 빅데이터의 속성에는 시계열 정보를 포함하고 있어야 한다.

따라서 본 연구에서는 공간정보 및 공간화 할 수 있는 빅데이터를 공간 빅데이터로 구분하기 보다는 빅데이터를 구성하는 데이터 유형 중 하나로써 위치정보와 해당 위치의 시계열 정보를 가지고 공간화 할 수 있는 모든 빅데이터를 공간 빅데이터로 정의할 수 있다.

그리고 선행연구에서의 빅데이터와 공간 빅데이터의 정의들을 검토한 결과 공간 빅데이터는 다른 유형의 빅데이터들과의 연계할 수 있는 분석기술과 활용체계까지 포함하여 고려해야 한다.

따라서 본 연구에서는 공간 빅데이터를 빅데이터를 구성하는 하나의 구성요소로써, 위치정보를 이용하여 공간화 할 수 있으며, 해당 위치의 시계열 정보를 가지고 공간화 할 수 있는 모든 빅데이터들과 이를 이용할 수 있는 활용체계를 공간 빅데이터라 정의하고자 한다.

따라서 Fig. 2와 같이 공간 빅데이터는 기존 빅데이터와 분리하여 구분할 것이 아니라, 기존 빅데이터를 구성하는 하나의 구성요소로서 이해하고, 이러한 활용체계 안에서 공간 빅데이터의 활용방안을 검토해야 한다.

3. 빅데이터 분석을 위한 공간 빅데이터 연계 서비스 방안

3.1 공간 빅데이터의 빅데이터 분석 서비스 요구 사항 분석

빅데이터의 정의를 다양한 정형, 반정형, 비정형 데이터들을 이용하여 가치있는 정보를 추출하고 활용할 수 있는 활용체계까지로 정의내리고 있었으며, 공간 빅데이터는 수많은 빅데이터의 유형 중 하나의 유형으로

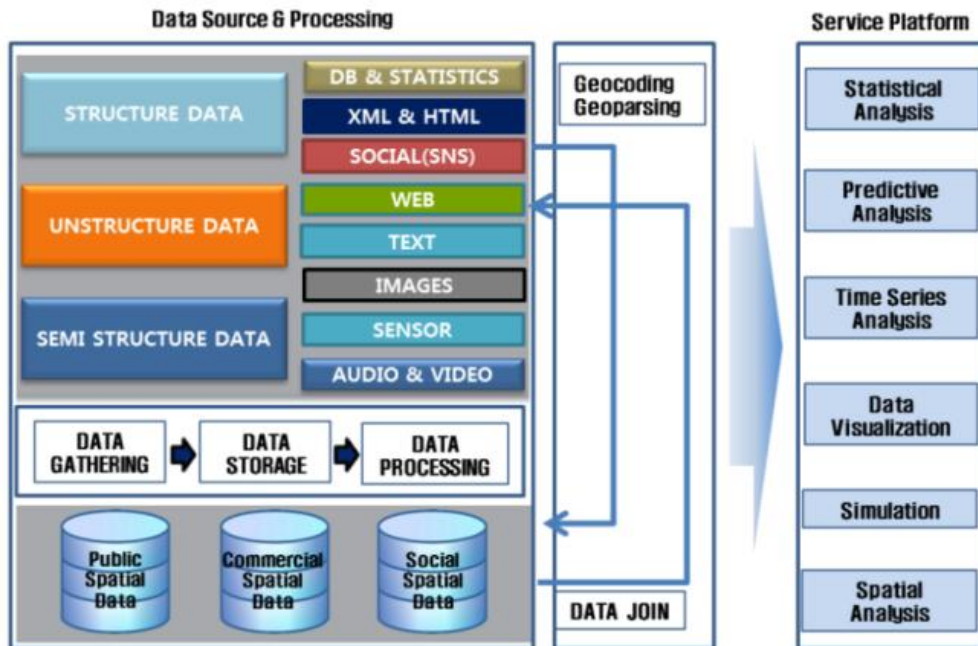


Figure 2. Big data analysis service system considering spatial big data

과약하였다. 따라서 Fig. 2처럼 공간정보를 포함한 공간 빅데이터와 다양한 빅데이터들과의 연계를 통한 공간분석 및 빅데이터 분석 서비스를 제공하기 위해서 공간 빅데이터는 다음과 같은 특징 및 요구사항을 제공해야 한다.

우선 공간정보를 포함한 공간 빅데이터는 기본적으로 다양한 공간분석이 가능해야 한다. 기본적인 중첩연산과 거리로 부터의 영향을 파악할 수 있는 근린연산, 공간적 분포와 영향을 알 수 있는 공간 상관분석 및 패턴분석과 핫스팟 분석 등 다양한 공간분석까지 분석할 수 있어야 한다.

또한 공간 빅데이터의 서비스 제공을 위해서 우선 공간정보 또한 지속적으로 구축 및 갱신되고 있어 기존에 구축된 공간정보와 향후 구축될 공간정보까지 고려할 수 있는 서비스 고려가 필요하다. 지속적으로 구축 및 갱신되고 있는 공간정보들은 사용자에게 구축 또는 갱신된 결과를 보여주고 있지만 지속적으로 누적된 공간정보를 이용한 분석결과를 이용한 다양한 서비스의 제공은 미흡한 실정이다. 따라서 기존 공간정보들을 이용한 공간분석 뿐만 아니라 향후 지속적으로 구축, 저장될 공간정보의 공간분석 및 빅데이터 분석까지 고려하여 서비스를 고려해야 한다.

그리고 시간의 흐름에 따른 위치별 속성정보의 시계열 변화를 탐지할 수 있어야 한다. 공간정보와 공간정

보화 가능한 빅데이터의 특징은 위치와 관련된 정보를 기반으로 공간정보가 지속적으로 갱신되고 있다면 해당 위치마다 시계열 변화에 따른 정보의 변화를 파악할 수 있어야 한다.

그리고 공간정보의 속성정보들을 이용하여 빅데이터 관련 분석이 가능해야 한다. 공간정보에 포함된 속성정보들을 이용하여 기존 빅데이터들과의 조인(join)을 통해 빅데이터 분석이 가능해야 한다. 기초적 통계량 분석부터 회귀분석까지의 기초 통계 분석과 속성 값들의 유사/상관분석 및 예측분석까지 가능해야 하며 분류, 추정, 예측, 군집화 및 유사집단화의 기능을 포함한 고차원의 데이터 마이닝분석과 이러한 분석 결과의 시각화도 가능해야 한다.

마지막으로 공간정보가 아닌 빅데이터들도 공간정보에 연계하여 공간 분석이 가능해야 한다. 이를 위해 빅데이터와 공간정보의 조인을 통한 공간정보화, 주소와 지점 등을 좌표로 공간화 하는 지오코딩(geocoding), 텍스트와 같은 비정형 데이터로부터 공간화 하는 지오파싱(geoparsing), 아이덴티티(identity)와 같이 해당 위치의 속성 값을 추출하는 중첩(overlay) 등을 이용하여 빅데이터와 공간정보를 연계하여 공간분석을 수행할 수 있어야 한다.

3.2 공간 빅데이터 분석 서비스 제공을 위한 격자 기반의 포인트 샘플링 방법

이러한 요구사항을 바탕으로 공간 빅데이터와 다양한 빅데이터들과의 연계를 통한 공간분석 및 빅데이터 분석 서비스를 제공하기 위해서는 격자 기반의 포인트 샘플링 방법을 제안하고자 한다. 위 방법은 Fig. 4와 같이 각각의 픽셀에 포인트를 샘플링 생성하여 Fig. 5와 같이 저장된 공간정보들 각각에 대해 해당 포인트 위치의 속성 값을 추출하여 다양한 분석에 활용하는 방법이다. 공간 분석 및 빅데이터 분석은 분석하고자 하는 이벤트가 발생했을 때 해당 이벤트 내용을 종속변수로 지정하고 포인트 샘플링을 통해 얻은 속성 값들을 이용하여 종속변수와 저장된 공간정보들의 속성정보들을 이용하여 분석에 활용하는 방법이다. 이러한 래스터 기반의 포인트 샘플링 방법을 이용하는 이유는 저장되는 공간정보들의 래스터 변환을 통해 포인트와 포인트들 사이에 일정한 간격으로 샘플 포인트들을 손쉽게 추출할 수 있을 뿐만 아니라, 회귀분석 및 상관분석과 같은 분석결과를 래스터로 변환된 공간정보들의 raster calculation을 통해 래스터 전체 격자에 쉽게 계산 및 적용 할 수 있다.

격자 기반 포인트 샘플링 방법은 크게 Fig. 3과 같이 가공 및 표준화를 통한 래스터기반의 저장, 분석을 위한 샘플링을 통한 속성정보 추출, 그리고 이러한 속성 값들을 이용하여 다양한 공간분석 및 빅데이터 기반의 분석을 수행할 수 있다.

우선 저장단계에서 기존에 구축되어 저장할 데이터와 향후 신규로 구축될 데이터들의 좌표체계 및 타원체를 표준화하고 동일 사이즈의 격자단위로 변경하여 정확히 같은 위치에서 속성 값들을 추출할 수 있도록 한다. 그리고 공간정보의 속성 하나당 하나의 파일로 저장될 수 있도록 공간 정보의 속성 값 별로 구분하여 공간화하여 저장한다. 이때 공간정보들은 공간DBMS기반의 공간 데이터 저장관리시스템에 저장하여 관리 할 수 있도록 한다.

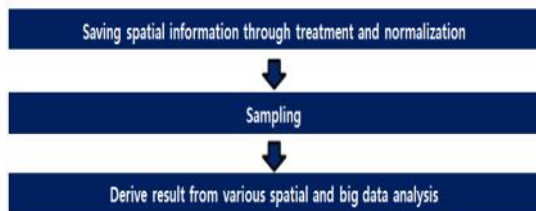


Figure 3. Flow chart of spatial big data analysis service

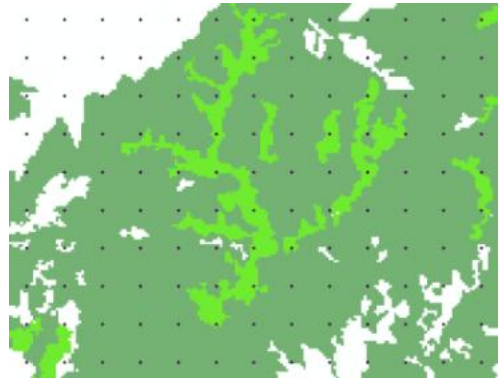


Figure 4. Example of grid-based point sampling

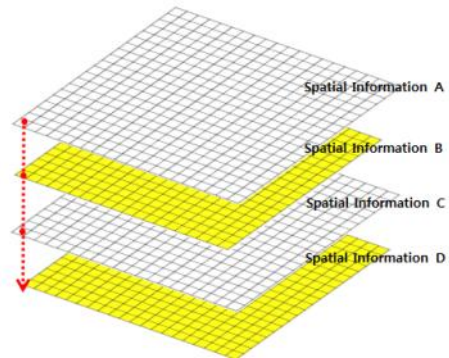


Figure 5. Extraction of grid-based point attribute value

그 다음 저장된 공간정보들의 규모에 따라 정의된 샘플링 포인트를 생성하고 Fig. 5와 같이 하나의 포인트를 기반으로 저장된 공간정보들의 속성정보를 추출한다. 이때 중요하게 고려해야 할 사항은 가로와 세로를 기준으로 해당 기준 격자 사이즈 기준으로 홀수의 곱만큼 떨어지도록 샘플링을 수행하도록 한다. 이를 위해 저장된 래스터의 리 샘플링을 통해 세분화된 격자를 하나의 격자로 생성한 후 래스터 별 포인트 생성을 진행하고, 이러한 포인트들을 이용하여 기존 래스터의 공간정보에 저장된 속성 정보들을 획득한다. 예를 들면 Fig. 6과 같이 10미터 단위격자로 공간정보들을 저장하였다면 샘플링을 수행할 때 동일 지점의 속성 값들을 추출하기 위해 10미터 범위로 샘플링 한다면 10미터 래스터의 중앙의 속성 값을 추출하고 상하 10미터의 간격을 두고 속성 값들을 추출하게 된다. 그리고 30미터의 범위로 속성 값을 추출할 경우에도 가로세로 전체 9개의 래스터 중 중앙의 값을 속성 값으로 획득하고 상하 30미터 떨어진 곳의 속성 값들을 획득하게 된다. 짝수의 곱으로 샘플링을 추출했을 때는 Fig. 6과 같이 기

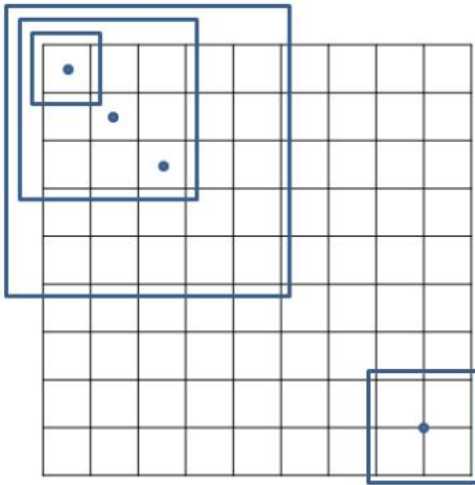


Figure 6. Example of creating sampling point and extracting attribute information using reference grid re-sampling

준 격자가 아닌 격자와 격자사이의 지점이 샘플링을 추출하게 되어 추후 속성 값을 이용한 분석이 어려울 수 있기 때문에 샘플링 시에는 홀수의 간격으로 샘플링한다.

이렇게 격자 기반의 포인트 샘플링 방법을 통해 저장된 공간정보들의 샘플링을 통한 속성 값들을 이용하여 어떠한 이벤트에 대한 분석들이 요구되었을 때, 앞에서 이야기 했던 공간 빅데이터 서비스를 위한 요구사항들을 만족하는 분석을 수행할 수 있다.

우선 샘플링 포인트에 저장된 속성정보와 포인트 위치정보를 이용하여 다양한 공간분석 수행이 가능하다. 예를 들어 우선 어떤 이벤트와 관련된 원인들을 찾고자 했을 때 기존에 발생했던 이벤트 결과들을 종속변수로 설정하고 이러한 종속변수의 공간정보 및 속성정보를 위 표준화된 공간정보 저장 체계에 일치시키고 해당 값들을 샘플링 포인트로 추출한다. 이렇게 추출된 종속변수는 포인트 기반으로써 기본적으로 공간 상관분석 및 패턴분석과 핫스팟 분석 등 공간분석이 가능하다. 물론 공간정보에 포함된 속성 값뿐만 아니라 다양한 빅데이터로부터 조인, 지오코딩, 지오파싱, 중첩을 통한 공간화(포인트화)를 통해 빅데이터로부터 다양한 공간분석을 수행할 수 있다.

그리고 이렇게 저장된 공간정보들은 데이터 생성 기준과 같은 시계열 정보를 속성 값으로 저장함으로써, 같은 위치에서의 변화된 시계열 정보들을 비교해 봄으로써 시간의 흐름과 관련된 다양한 시계열 분석을 수행

할 수 있고, 향후 미래의 예측과 관련된 분석들을 수행할 수 있다. 그리고 향후 지속적으로 구축 될 공간정보들의 가공, 표준화, 저장을 수행하고 해당 위치별 속성 정보를 추출하는 자동화된 과정을 통해 다양한 시계열 분석을 수행할 수도 있다.

그리고 기존에 저장된 공간정보들간의 속성 정보(값)들과 빅데이터로부터 추출한 정형화된 데이터들의 속성 값과의 조인을 통해 상관분석 및 다양한 빅데이터 분석이 가능하다. 또한 기존의 공간정보에 구축 시점과 같은 시간과 관련된 위치 값과 속성 값들을 이용하여 위치정보와 속성정보의 변화를 탐지하고 이를 기반으로 향후 미래의 예측도 가능할 것이라고 판단된다.

이렇게 기존 공간정보와 연계한 빅데이터 분석을 통해 기존에 예측하지 못했던 새로운 가치 있는 연구결과를 통해 현안문제 진단 및 해결을 위한 정책지원 및 대국민 서비스 지원이 가능할 것으로 판단된다. 그리고 이러한 공간정보의 빅데이터의 활용은 공간정보의 활용이 증대될 것이며, 공간정보의 활용 증대는 관련 공간정보 분야의 발전에도 큰 기여를 할 수 있을 것이라 판단된다.

4. 결 론

본 연구에서는 빅데이터와 공간 빅데이터의 연구 사례들을 분석하여 빅데이터의 정의와 공간 빅데이터의 정의 및 특징들을 기반으로 공간 빅데이터의 정의를 내렸다.

공간 빅데이터를 빅데이터를 구성하는 하나의 데이터 유형 중 하나의 구성요소로써, 위치정보를 이용하여 공간화 할 수 있으며, 해당 위치의 시계열 정보를 가지고 공간화 할 수 있는 모든 빅데이터들과 이러한 데이터들을 이용할 수 있는 활용체계를 공간 빅데이터라 정의하였다.

그리고 이러한 정의의 기반으로 공간 빅데이터가 제공해야 하는 기본적인 공간 빅데이터 서비스들을 제시하였다. 그리고 이러한 요구사항들을 만족시킬 수 있는 공간 빅데이터 서비스 제공을 위해 샘플링 포인트 생성 및 속성정보 추출 방안을 제시하였다.

샘플링 포인트 생성 및 속성정보 추출 방안은 공간 빅데이터가 제공해야 하는 다양한 공간분석이 가능하며, 해당 위치 별 시간의 변화에 따른 정보의 변화를 파악할 수 있고, 향후 구축될 공간정보까지 고려하여 분석에 이용할 수 있다. 그리고 공간정보의 속성정보들을 이용하여 다양한 빅데이터 관련 분석이 가능하며, 공간 정보가 아닌 다양한 형태의 빅데이터들도 가공을 통해

공간정보에 연계하여 공간 분석에 이용할 수 있다.

이러한 공간정보들과 기존 빅데이터를 하나의 빅데이터의 구성요소로 인식하고 다양한 빅데이터간의 결합 또는 융합을 통한 다양한 분석에서의 활용은 우리가 기존에 예측했던 결과 이외의 다양하고 가치있는 결과를 도출할 수 있으며, 이러한 분석결과는 효율적 정책적 집행과 지원이 가능할 수 있으며 만족도 높은 대국민 서비스를 지원할 수 있다. 그리고 이러한 빅데이터와 연계된 공간정보의 활용 증대는 공간정보 산업 및 기술발전에 크게 기여할 수 있을 것이라 판단된다.

본 연구는 공간 빅데이터 서비스 제공을 위한 개념적 연구라는 한계가 있지만, 향후 본 연구에서 제시한 격자 기반의 포인트 샘플링 방법론을 적용하여 빅데이터와 공간 빅데이터의 연계를 반영할 수 있는 자동화된 공간 빅데이터 관리 및 분석 시스템 구축에 충분히 활용 가능할 것이라는 점에서 의의가 있다고 판단된다.

감사의 글

본 연구는 한국환경정책·평가연구원(KEI)에서 수행된 ‘북한 환경정보 구축 및 활용 방안 연구 2 : 원격탐사를 이용한 자연환경성 우수지역 평가와 환경부 사업으로서 한국환경정책·평가연구원(KEI)에서 수행한 “2016년 국가환경지도 구축운영 사업” 과제의 지원에 의하여 수행되었습니다.

References

- Ahn, J. W., Yi, M. S and Shin, D. B., 2013, Study for spatial big data concept and system building, Journal of Korea Spatial Information Society, Vol. 21, No. 5, pp. 43-51.
- Beyer, M. A. and Laney, D., 2012, The importance of big data: a definition, Technical report, Gartner, USA, pp. 3-4.
- Dijcks, J. P., 2014, Oracle: big data for the enterprise, Technical report, Oracle, USA, pp. 3-4.
- Gantz, J. and Reinsel, D., 2011, Extracting value from Chaos, Technical report, IDC, USA, pp. 1-12.
- James, M. and Michael, C., 2011, Big data : the next frontier for innovation, competition, and productivity, Technical report, McKinseyGlobal Institute, USA, p.1.
- Kim, D. H. et al., 2014, The use of spatial big data for planning policy support, Research report, KRIHS, Republic of Korea, pp. 13-26.
- Kim, D. J. and Yoon, S. Y., 2013, Big data utilization for monitoring territorial policy responses and predicting policy demand, Research report, KRIHS, Republic of Korea, pp. 21-69.
- Kim, D. J., Hwang, M. H., Yoon, S. Y. and Seo, T. S., 2014, A study on building and utilization of spatial knowledge platform for scientific territorial planning(1), Research report, KRIHS, Republic of Korea, pp. 1-49.
- Kim, H. S., 2014, Policy direction to the establishment and application of spatial big data system, Planning and Policy, Vol. 389, No. 4, pp. 6-11.
- Kim, M. J., Kim, D. J. and Lee, Y. J., 2013, Spatial big data utilization for national land policy, Research report, KRIHS, Republic of Korea, pp. 21-69.
- Lee, Y. J., Kim, M. J. and Em, E. S., 2014, Implementation of welfare and safety national land and application of spatial big data, Planning and Policy, Vol. 389, No. 4, pp. 12-21.
- National Council on Informatization Strategies, 2011, Implementation of smart government using big data, Research report, National Council on Informatization Strategies, Republic of Korea, pp. 1-27.
- Park, S. M et al., 2013, The new competitiveness of firms big data curation, Technical report, SERI, Republic of Korea, pp. 1-17.
- Teradata, 2013, Teradata unified data architecture, Technical report, Teradata, USA, pp. 1-6.
- Yi, M. S., Lee, C. H. and Kim, J. Y., 2014, Big data analysis on demands for environmental policies, Research report, KEI, Republic of Korea, pp. 5-9.
- Yu, S. C., Choi, W. W., Shin, D. B. and Ahn, J. W., 2014, A Study on concept and services framework of geo-spatial big data, Journal of Korea Spatial Information Society, Vol. 22, No. 6, pp. 13-21.