

## 공공 빅데이터의 시각화를 위한 InfograaS의 아이디어 제안

### Idea proposal of InfograaS for Visualization of Public Big-data

차병래<sup>1</sup> · 이형호<sup>2</sup> · 심수정<sup>3</sup> · 김종원<sup>1\*</sup>

<sup>1</sup>광주과학기술원 정보통신공학부

<sup>2</sup>가민정보시스템 전략사업본부

<sup>3</sup>전남대학교 전산학과

Byung-Rae Cha<sup>1</sup> · Hyung-Ho Lee<sup>2</sup> · Su-Jeong Sim<sup>3</sup> · Jong-Won Kim<sup>1\*</sup>

<sup>1</sup>School of Information and Communications, GIST, Gwangju 500-712, Korea

<sup>2</sup>Headquarters for business strategy, Comin Information System Inc, Gwangju 503-841, Korea

<sup>3</sup>Department of Computer and Science, Chonnam National University, Gwangju 500-757, Korea

#### [요 약]

본 논문에서는 클라우드 컴퓨팅 자원을 이용하여 빅데이터의 일종인 LOD (linked open data)를 가공 및 분석하는 방법을 제안한다. LOD는 공공 데이터를 공유 및 재활용하기 위한 웹기반의 오픈 데이터이다. 특히 BA(business analytics)와 Info-graphic을 위한 시각화 (visualization) 기술을 제공하는 새로운 SaaS (software as a service) 비즈니스 영역을 InfograaS (Info-graphic as a service)라고 정의한다. 본 연구의 목표는 시각화 및 비즈니스 전문가 없이 비전문가 또는 초보자가 사용할 수 있도록 하는 것이다. 데이터 시각화 (data visualization)는 데이터 분석 결과를 쉽게 이해할 수 있도록 시각적으로 표현하고 전달되는 과정을 말한다. 데이터 시각화의 목적은 차트와 그래프를 통해 정보를 명확하고 효과적으로 전달하는 것이다. 공공기관의 빅데이터를 클라우드 컴퓨팅 자원과 오픈 소스인 하둡, R, 기계학습, 데이터 마이닝 등을 이용하여 다양한 처리 결과를 이해하기 쉬운 그래픽 또는 차트로 표현하고 공유한다.

#### Abstract

In this paper, we have proposed the processing and analyzing the linked open data (LOD), a kind of big-data, using resources of cloud computing. The LOD is web-based open data in order to share and recycle of public data. Specially, we defined the InfograaS (Info-graphic as a service), new business area of SaaS (software as a service), to support visualization technique for BA (business analytics) and Info-graphic. The goal of this study is easily to use it by the non-specialist and beginner without experts of visualization and business analysis. Data visualization is the process to represent visually and understand the data analysis easily. The purpose of data visualization is to deliver information clearly and effectively by chart and figure. The big data of public data are shared and presented in the charts and the graphics understood easily by various processing results using Hadoop, R, machine learning, and data mining of open source and resources of cloud computing.

**Key word:** Big-data, Linked open data, InfograaS, Business analytics, Info-Graphic.

<http://dx.doi.org/10.12673/jant.2014.18.5.524>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 1 September 2014; Revised 28 October 2014

Accepted (Publication) 6 October 2014 (30 October 2014)

\*Corresponding Author; Jong-Won Kim

Tel: +82-62-715-3136

E-mail: jongwon@gist.ac.kr

## 1. 서론

빅데이터(big data)란, 인터넷, 카카오톡, 페이스북, 트위터 등을 통해 오가는 모든 메시지, 이미지, 그리고 영상 등을 포괄하는 용어를 의미한다. SNS(social network service)뿐만 아니라 GPS(global positioning system)를 기반으로 한 지도 정보, 날씨 정보처럼 현존하는 정보들을 ‘중요한 데이터’라는 개념으로 정의하고 주목하기 시작한 데서 ‘빅데이터’의 시대가 시작됐다. 또한, 소셜 네트워크 서비스가 활성화되기 시작하면서 활자뿐만 아니라, 업데이트 되는 데이터의 90%가 이미지, 동영상 등 다양한 형태를 보이고 있다. 이처럼 빅데이터는 지금까지 없었던 것이 아니라, 우리가 지금까지 인식하지 못했을 뿐이며, 빅데이터이라는 새로운 이름으로 재정의된 것이다. 또한 최근 많은 기업들이 빅데이터로 사람의 욕망을 읽어낼 수 있다는 데에 큰 매력을 느끼기 시작했으며, 이를 통해 마케팅 도구로 활용할 수 있다고 많은 기대를 하고 있다.

빅데이터 분석을 통한 지능형 교통정보서비스 사례는 스웨덴의 로얄 기술 대학(royal institute of technology)는 운행 중인 자동차들의 GPS 데이터, 도로의 레이더 센서를 통한 교통 유량과 운행차량의 속도 데이터, 교통 혼잡 부담금 데이터, 기후 정보, 교통사고 및 도로 공사 정보를 토대로 실시간 교통정보를 제공하고 있다. 모든 데이터는 대용량의 정형, 비정형 데이터를 리얼타임으로 분석해냄으로써 현 교통상황 및 목적지까지의 경로별 소요 시간을 계산해 요청자에게 제공하고 있다. 또한 일본은 노무라연구소와 건설성, 통신성, 운수성, 우정성, 경찰청 5개 정부 기관이 협력하여 지능형 교통안내 시스템(intelligent transport system) 서비스를 제공하고 있다. 일본 전역의 택시 11,000대와 교통정보 제공에 동의한 일반 운전자들의 GPS를 통해 자동차의 주행 스피드를 계산하는 방식으로 실시간 교통정보를 수집하고 이를 사용자들의 스마트폰으로 송신한다. 2011년 이와테현의 대지진이 발생했을 때 일본의 교통안내 시스템은 구조 차량 및 지원자원 수송 차량이 지진 발생 지역의 교통 체증을 피해 신속하게 현장에 접근할 수 있게 함으로써 지진 피해를 최소화시키는 데 일조했다는 평가를 받고 있다. 또한, 많은 빅데이터 사례들 중에서 서울시의 올빼미 버스[1]는 우리나라에서의 공공부문 빅데이터 사업의 우수 사례로 꼽히고 있다.

본 논문에서 제안하는 InfograaS(Info-graphic as a service)는 SaaS의 일종이며, 그림 1과 같이 중소기업 및 비전문가들에게 빅데이터 처리 결과를 이용하여 BA (business analytics) 분석과 웹 기반의 시각화가 가능하게 하며, 그림 2와 같이 TDWI research에서는 빅데

이터 분석 시장의 잠재 성장성을 높게 평가하고 있다 [2]. 제안하는 InfograaS의 중요성으로는 빅데이터 분야의 초기 분석 단계로써 원 데이터를 분석하여 BA 분석을 위한 사전 정보를 획득할 수 있으며, 이를 기반으로 비전문가도 BA의 다양한 툴을 선택하여 다양한 시각 측면에서의 정보를 제공하게 된다. 더불어 이러한 정보는 Info-Graphic의 다양한 시각화 툴들을 제공하여 빅데이터 안에 존재하는 해안을 제공하게 된다. 또한, InfograaS는 다양한 BA와 시각화 툴, 그리고 빅데이터 처리가 필요한 다양한 새로운 영역에서도 사용될 수 있다.



그림 1. 빅데이터의 시각화

Fig. 1. Visualization of big-data.

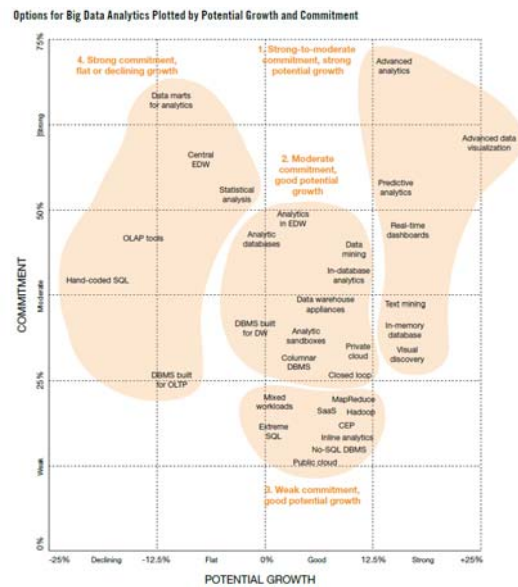


그림 2. 빅데이터 분석의 잠재 성장성

Fig. 2. Potential growth of big-dta analysis.

## II. 관련 연구

### 2-1 미국의 공공부문 빅데이터 활용 동향

미국 오바마 정부는 ‘Open Government Initiative’를 기반으로 중앙정부 및 연방정부의 공공정부에 대한 일반 시민들의 접근성을 강화하고 상업적 활용을 촉진하는 방향으로 지원하고 있다. Open Gov를 통해 모든 연방정부의 행정 데이터를 공개하도록 지시하고 현재 172개 기관에서 보유한 390,178건의 data set을 제공 중(‘12.1 기준)이다. 특히, 데이터의 표준화를 위해 공통된 API를 적용해 데이터를 제공하고 있으며, 데이터 자체 뿐만 아니라 1차 분석된 시각화된 자료로 함께 제공하고 있다. 미국 Socrata사는 공공데이터를 개방하고자 하는 연방정부, 공공기관을 대상으로 공통된 플랫폼을 구축해주는 사업 추진 중이며, 시카고, 시애틀, 오레곤, 워싱턴 등 지방정부가 Socrata의 플랫폼을 이용하고 있으며, 표준화된 데이터 세트를 제공하기 위한 API, 클라우드 시스템, 시각화 시스템 등을 제공하고 있다.

### 2-2 공공부문 빅데이터 활용 동향

우리나라는 ‘공공기관 정보공개에 관한 법’을 통해 비공개 대상정보를 제외하고 모든 공공데이터는 공개하고 있으며, ‘11.7월 이후 공유자원포털을 통해 원문 데이터 7만 여건을 개방하여 무료로 사용할 수 있다. 기상정보, 공간정보 등 민간수요가 높은 일부 공공정보는 개별법에 제공 근거를 마련하고 민간에 제공되고 있으며, 표 1은 공공기관별 공공정보 개방 현황을 나타낸 것이다.

표 4. 기관별 주요 공공정보 개방 현황

Table. 1. Open state of major public data by government system.

구분	제공 정보
국토해양부	공간정보(국가공간정보에 관한 법률), 부동산 정보 등
특허청	산업재산권정보(발명진흥법), 전문가 특허정보(약 20종) 등
농촌진흥청	농업정보서비스(농업기술, 병충해, 작물 정보, 품종정보, 가격, 유통 등)
통계청	일반통계정보(통계법), 공간통계 등
문화체육관광부	한국전통문양 등
서울시	서울 열린 데이터 광장(‘11.2 개설)을 통해 ‘14년까지 150종 개방

그러나 현재 공개되고 있는 대부분의 데이터는 가치 있는 원시데이터와는 거리가 멀고, 실제로 검색을 통한 접근은 불가능하다. 통계청에서 제공하는 수치로 표시된 데이터를 제외하고 대부분의 데이터는 문서 자료로 이를 활용해 데이터를 추출하거나 분석하기에는 무리가 있다. 이제는 그림 3과 같이 실제 데이터로써 활용 가능한 표준화된 데이터 세트를 공개할 필요가 있다.

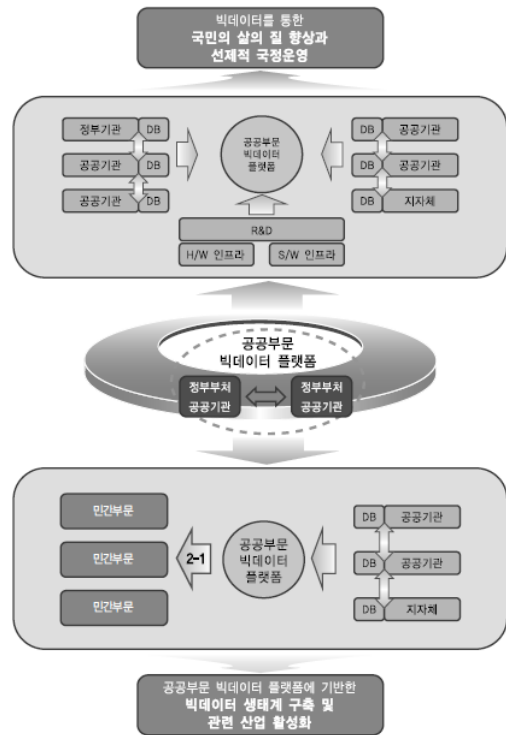


그림 3. 공공부문 빅데이터 추진 방향

Fig. 3. Drive direction of Public big-data.

2014년 4월 공공데이터의 데이터셋 활용신청 수

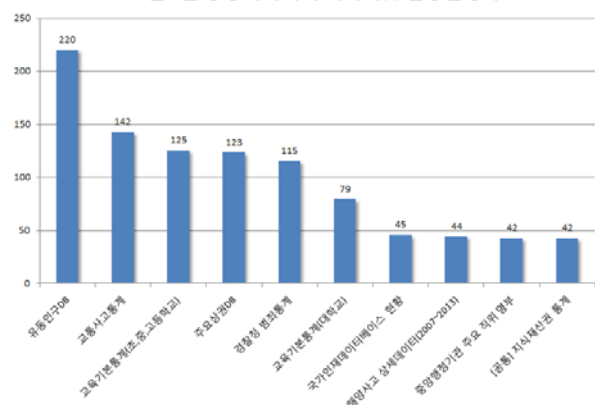


그림 4. 공공데이터포털의 데이터셋 활용신청 수

Fig. 4. Datasets utilization number of public data portal.

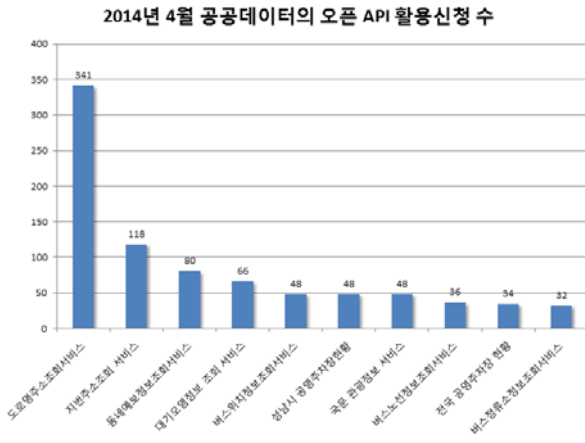


그림 5. 공공데이터포털의 Open API 활용신청 수  
Fig. 5. Open API utilization number of public data portal.

그림 4와 그림 5는 2014년 4월에 사용된 공공데이터 포털의 데이터셋 활용 수와 OpenAPI 활용 수를 나타낸 것이다.

### 2-3. 빅데이터의 국내외 시장 현황

빅데이터의 활용분야가 넓어짐에 따라 빅데이터의 시장 전망도 점점 긍정적으로 변화하고 있다. 리서치 및 분석 업체 Wikibon\*에 따르면, 2014년 올해는 168억 달러를 돌파할 것으로 예상하고 있다. 또 2015년에는 올해의 수입을 뛰어넘어 거의 2배의 성장률을 보이고 있으며, 그만큼 빅데이터의 사용이 대중화되고, 각종 세계 시장에서 유용하게 활용될 것으로 예측할 수 있다.

국내 ICT 분야의 시장 규모는 정보통신산업진흥원이 제공하는 ITSTAT 서비스를 이용하였으며, 빅데이터 산업의 주요 5개 분야를 대상으로 조사하였다. 그림 7은

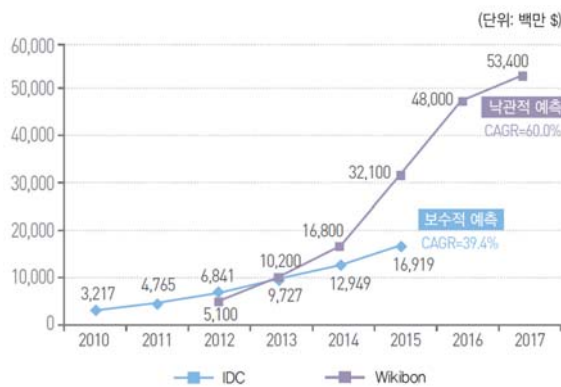
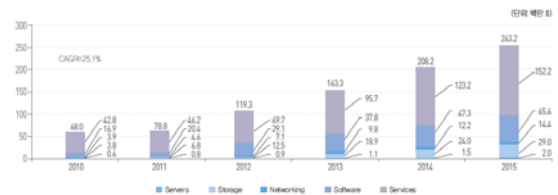


그림 6. KISTI의 빅데이터 산업의 현황과 전망  
Fig. 6. Present state and prospect of big-data industry by KISTI.

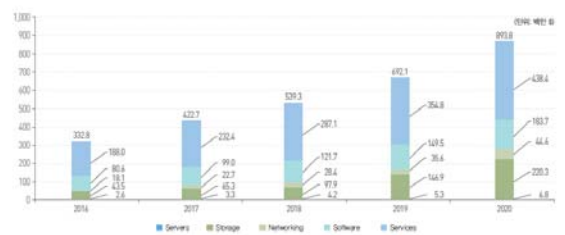
각각의 5개 분야별 국내 시장 규모를 나타낸 것이다. 해당 분야의 2013년 이후 시장 성장률은 최근 3년간 (2010~2012년)의 평균 성장률을 적용하였으며, 국내 빅데이터 시장 규모의 산출 결과로는 단기적으로 2015년 약 263억만 달러(한화 3천억 원)에 이를 것으로 예상되며 세계 빅데이터 시장의 약 1.6%의 비중을 점유할 것으로 전망되었다. 중장기적으로는 국내 빅데이터 시장 규모가 2020년에는 약 900억만 달러(한화 1조 원)에 이를 것으로 예상되었다. 이러한 추세로 국내 빅데이터 시장이 성장할 경우에는 국내 ICT 분야의 관련 산업에서 빅데이터 분야가 차지하는 비중은 2013년 0.6%에서 지속적으로 증가하여 2020년에는 약 2.6%에 이를 것으로 전망되었다.



단위: 백만 \$

구분	2010년	2011년	2012년	2013년	2014년	2015년	CAGR
Servers	0.6	0.8	0.9	1.1	1.5	2.0	27.8%
Storage	3.8	6.8	12.5	18.9	24.0	29.0	50.0%
Networking	3.9	4.6	7.1	9.8	12.2	14.4	25.3%
Software	16.9	20.4	29.1	37.8	47.3	66.6	22.9%
Services	42.8	46.2	69.7	95.7	123.2	152.2	23.6%
합계	68.0	78.8	119.3	163.3	208.2	263.2	25.1%
세계 시장 비중	2.1	1.7	1.7	1.7	1.6	1.6	-

그림 7. 국내 빅데이터 시장 규모의 단기 전망  
Fig. 7. Short-term prospect of domestic big-data market scale.



단위: 백만 \$

구분	2010년	2011년	2012년	2013년	2014년	2015년	2020년
Servers	2.6	3.3	4.2	5.3	6.8	8.8	13.7
Storage	43.5	65.3	97.9	146.9	220.3	334.4	520.3
Networking	18.1	22.7	28.4	35.6	44.6	55.3	84.6
Software	80.6	99.0	121.7	149.5	183.7	226.3	343.7
Services	188.0	232.4	287.1	354.8	436.4	536.4	816.4
합계	332.8	422.7	539.3	692.1	892.1	1,113.8	1,784.8

그림 8. 국내 빅데이터 시장 규모의 장기 전망  
Fig. 8. Long-term prospect of domestic big-data market scale.



2-4 BI의 기술 동향

가트너에 따르면 기업의 비즈니스 인텔리전스(BI; business intelligence) 기능을 확장하기 위해 클라우드 서비스 혹은 SaaS를 이용하는 사례가 증가할 것으로 전망되었다. 가트너 그룹이 2011년 4분기에 BI 플랫폼을 사용하는 1,364명의 IT 관리자와 현업 사용자를 대상으로 실시한 조사 결과에 따르면, 17%의 기업이 핵심 BI 기능의 일부를 클라우드 서비스 및 SaaS 기반으로 교체하였거나 교체할 계획이었다. 한편, 핵심 BI 기능이 아닌 특정 사업 분야 및 안전에 대한 BI 기능을 확장하기 위해 클라우드 서비스를 이미 사용하고 있거나 향후 12개월 내에 사용을 계획 중인 기업은 27%이었다. BI 기능을 원하는 현업 사용자들은 종종 배포 주기, 비용, 복잡한 업그레이드 프로세스, BI 솔루션이 요구하는 IT 인프라 조건 등의 요인으로 인해 어려움을 겪고 있는 상황이었다. 이러한 상황 때문에 SaaS와 클라우드 기반 서비스들은 엔터프라이즈 BI 플랫폼과 기업 성능관리 제품 시장을 교란하기 보다는 아직까지 BI에 투자하지 못하는 중소기업 시장이나 특정 영역에 대한 추가적인 분석 기능을 제공하는 등 새로운 기회를 만들어 내는 측면이 강하다. 그러나 가트너 그룹은 클라우드 서비스 및 SaaS 기반의 BI와 분석의 정확한 의미가 무엇인지, 이를 통해 무엇을 얻을 수 있는지에 대한 명확한 시장의 합의는 아직 정립되지 않았다고 지적하고 있다.

클라우드 서비스 및 SaaS 기반의 BI를 이용하거나 이용을 고려중인 기업은 클라우드 서비스의 이용 비용 및 위험에 대해 지속적으로 점검하고 대응하는 전략의 실행이 필요하다. 이미 비즈니스 운영에 필요한 애플리케이션들을 클라우드 서비스 및 SaaS 형태로 이용하는 기업들이라면, 클라우드 기반 BI 및 분석 서비스 이용을 우선적으로 고려할 수 있다. 그러나 클라우드 서비스 및 SaaS 기반의 BI를 이용하는 기업들은 반드시 항시적인 위험을 평가해야 하며, 자신들이 선택한 클라우드 제공업체가 실행 가능한 결과를 제공할 수 있는

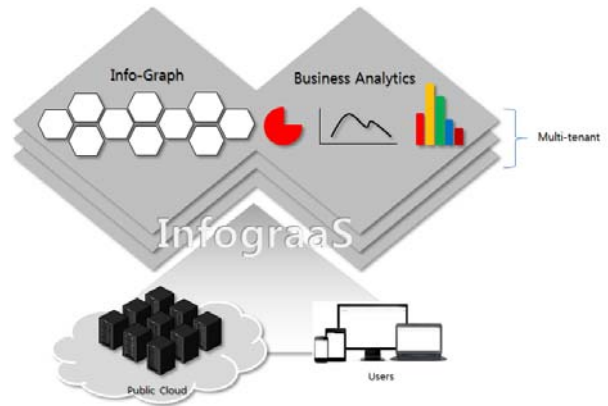


그림 10. InfograaS의 개념도  
Fig. 10. Concept diagram of InfograaS.

적합한 비즈니스 역량을 제공할 수 있는지 확인이 필요하다.

또한, 클라우드 서비스를 이용하는 기업들은 클라우드 분석이 또 하나의 폐쇄적인 채널이 되지 않도록 분석 데이터가 다른 솔루션들 사이에서 자유롭게 흘러다닐 수 있도록 기업의 BI 전략을 확실히 정립해야 한다. 그림 9의 2011년 9월 발표된 가트너 그룹의 하이프 사이클 예측에 따르면 SaaS 기반 비즈니스 인텔리전스는 2013년 2016년 사이에 안정적인 궤도에 접어들 것으로 전망하고 있다.

III. InfograaS의 정의 및 설계

퍼블릭 클라우드 컴퓨팅의 자원을 이용하여 공공기관의 빅데이터 ([https:// www.data.go.kr](https://www.data.go.kr), 광역시)를 오픈소스 도구로 가공 및 분석하여 정보 수요자에게 분석 데이터인 BA와 시각화를 지원하기 위한 Infor-Graphic을 제공할 수 있는 SaaS의 한 종류인 InfograaS를 설계/개발하고 정보들을 공유한다. 아래 그림 10과 같이 나타낼 수 있다.

이를 위하여 그림 10과 같이 크게 4개 부문으로 구분할 수 있으며, 공공기관 및 지자체의 빅데이터를 제공하는 공공데이터포털 부문, 빅데이터를 가공 및 분석하기 위한 컴퓨팅과 스토리지의 자원을 제공하는 인프라 분야, 빅데이터와 컴퓨팅 및 스토리지 리소스를 이용하여 오픈소스 기반의 SaaS 분야, 그리고 빅데이터에 존재하는 해안을 찾기 위한 BA와 Infographic 분야로 구성된다.

3-1 InfograaS의 기능

InfograaS의 기능은 다음의 그림 11과 같이 5가지 기능들로 정의할 수 있다.

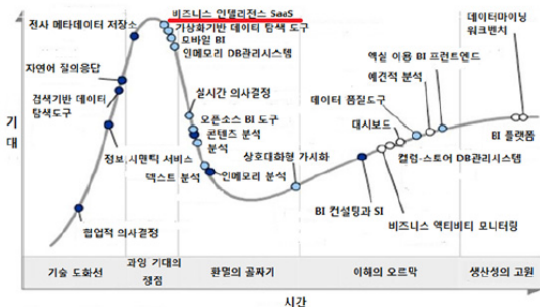


그림 9. 비즈니스 인텔리전스 관련 기술들의 하이프 사이클  
Fig. 9. Hype-cycle of business intelligence related techniques.



그림 11. 빅데이터 분석의 프레임워크  
 Fig. 11. Framework of Big-data analysis.

- 공공 데이터 및 LOD의 빅데이터 수집 및 저장 기술
- 퍼블릭 클라우드의 리소스 할당과 백업을 위한 클라우드 버스팅과 스페닝 기술
- 빅데이터 분석을 위한 하둡 기술
- BA를 위한 통계 및 기계학습, 데이터 마이닝 기술
- 시각화 기술

최근 빅데이터 중에서 가장 가치 있는 데이터로는 공공 데이터이며, 공공 데이터 및 LOD의 빅데이터 수집 및 저장 기술은 다양한 서비스를 제공하기 위한 기본 절차이다. 공공 데이터를 수집한 후에 프라이빗 클라우드 또는 퍼블릭 클라우드의 자원을 이용하여 공공 데이터를 가공 및 분석하고 퍼블릭 클라우드의 리소스 할당을 위해서는 클라우드의 버스팅 기술과 스페닝 기술[3]이 필수적이다. 클라이언트 또는 프라이빗 클라우드에서 빅데이터의 가공이 불가능하면, 클라우드 스페닝 기술에 의한 자원을 프로비저닝이 가능하다. 빅데이터 분석을 위한 하둡 기술은 퍼블릭 클라우드에 가상 머신들을 생성하고, 이 가상 머신들에 오픈소스 하둡을 설치하며, 설치된 하둡을 이용하여 빅데이터의 분산 병렬처리를 수행한다.

최근 동향으로 BI에서 점차적으로 BA로 진화하고 있는 상황이며, BA를 위한 통계 및 기계학습, 그리고 데이터 마이닝 기술에 의해서 빅데이터 안에 존재하는 해안을 얻을 수 있게 된다. 시각화 기술은 공공 정보 LOD 등의 수집된 빅데이터를 하둡, 통계, 기계학습, 데이터마이닝 등의 다양한 처리 결과를 이해하기 쉽게 다양한 그래픽 또는 차트로 표현하게 된다. 빅데이터 처리 결과를 다양한 플랫폼에서 지원할 수 있도록 익스플로러, 크롬 등의 웹 브라우저 기반으로 구현할 예정이며, 빅데이터 처리 결과인 그래픽 또는 차트가 일방적인 표현이 아니라, 단말 사용자와 인터랙티브하게 반응하도록 시각화 기술을 개발할 예정이다.

3-2 Info-graphic의 샘플들

성별에 따른 년도별 취업자 수

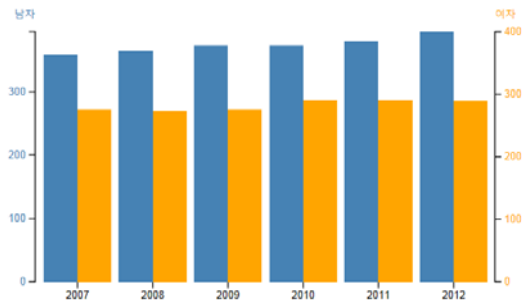


그림 12. 광주광역시의 성별에 따른 년차별 취업자 수  
 Fig. 12. Employment number about distinction of sex per year in Gwangju metro-city.

2007 ~ 2012년의 광주광역시 북구 지역의 외국인 등록자 수

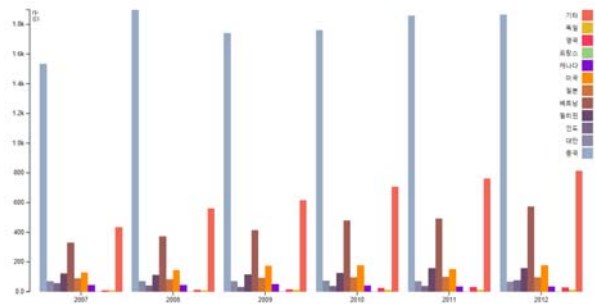


그림 13. 광주광역시 북구 지역의 외국인 등록자 수  
 Fig. 13. Foreigner registration number on buk-gu, Gwangju metro-city.

공공 데이터 포털[4]의 일부분을 가공하여 인포그래픽들을 생성하였으며, 그림 12는 D3(data-driven document)[5]에 의한 광주광역시의 성별에 따른 년도별 취업자 수를 나타내는 차트이며, 그림 13은 년차별 광주광역시 북구 지역의 외국인 등록자 수를 나타내는 바 차트이다. 그리고 그림 14는 광주광역시의 자동차 등록 건수를 파이 차트로 나타낸 것이다.

그림 15는 R(<http://www.r-project.org>)을 이용하여 최근 약 50년 동안 미국의 실업률을 분기별 변화의 증가 및 감소 변화를 나타내는 그래프이다. 이 그래프에서는 윌리엄 클리블랜드와 수잔 데블린은 LOESS(locally weighted scatterplot smoothing)이라는 이름의 통계적 방법론으로 데이터의 곡률에 맞는 추세선을 나타내고 있다. 그림 16은 미국의 주별 범죄율을 스타 차트로 나타낸 것이다. 스타 차트는 몇 개의 축을 그리고, 전체 공간에서 하나의 변수마다 축 위의 중앙으로부터의 거리로 수치를 나타낸다. 중점은 축이 나타내는 값의 최소값을, 가장 먼 끝은 최대값을 나타낸다. 하나의 대상

### 2007 ~ 2012년의 자동차 등록 건수

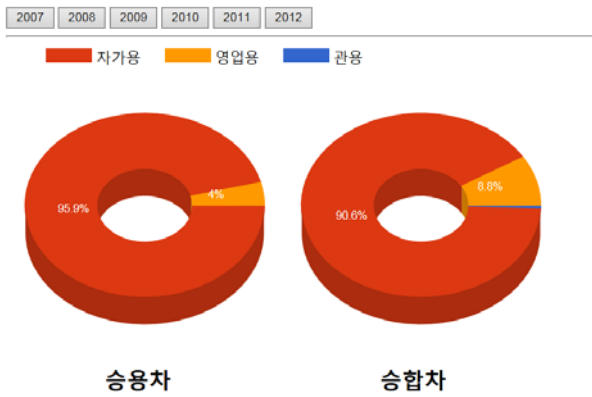


그림 14. 광주광역시의 2007 ~ 2012년의 자동차 등록 건수  
**Fig. 14.** Auto-moblie registration number between 2007 and 2012 in Gwangju metro-city.

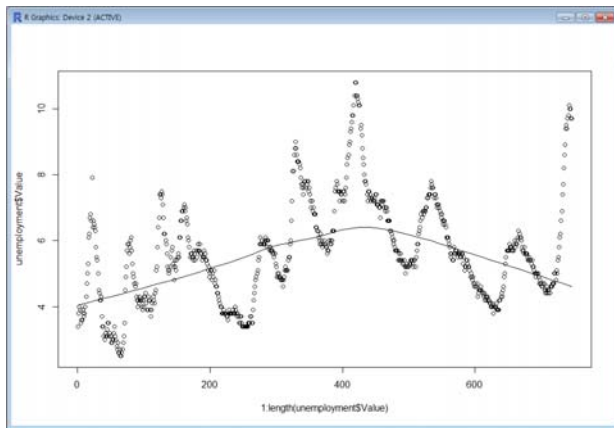


그림 15. 미국의 약 50년간의 실업률의 분기별 변화  
**Fig. 15.** Unemployment-rate during a quarter for 50 years year in USA.

에 대한 스타 차트를 그리면, 하나의 변수에서 다른 변수로 이어지는 연결선을 그린다. 그 결과는 별모양의 도형으로 나타나게 된다. 그림 17은 세계 지도위에 출산율을 버블 차트로 나타낸 것이며, 버블 차트는 반지름이 아닌 면적으로 양을 표시 한다.

#### IV. 결론

빅데이터의 성장에 따른 빅데이터 분석 시장의 잠재 성장성이 증가하고 있다. 클라우드 컴퓨팅 자원을 이용하여 빅데이터의 일종인 LOD를 가공 및 데이터 분석을 위한 InfograaS를 제안하였다. 제안된 InfograaS는 중소기업 및 비전문가들에게 빅데이터 분석 결과를 BA 분석 결과와 웹 기반의 시각

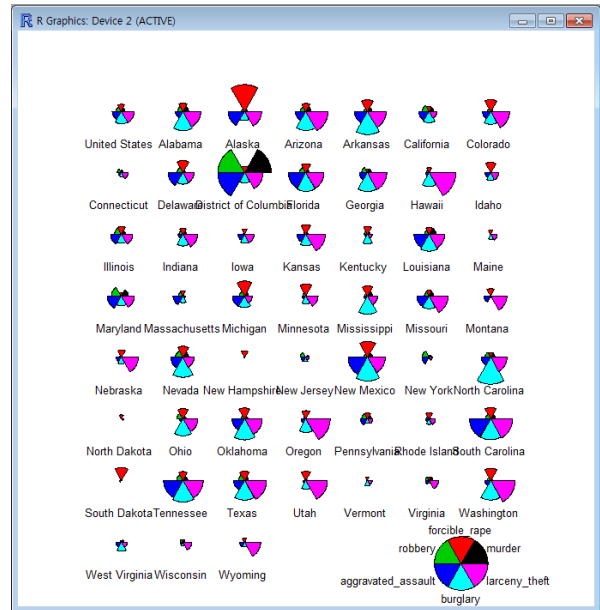


그림 16. 미국의 주별 범죄율  
**Fig. 16.** Crime-rate of states in USA.

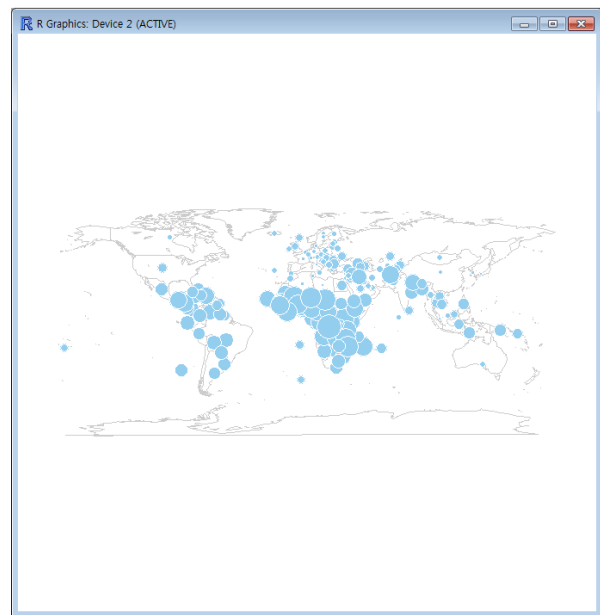


그림 17. 세계 출산율의 버블 차트  
**Fig. 17.** Bubble chart of world birth-rate.

화로 지원이 가능하게 하며, 제안하는 InfograaS의 중요성으로는 빅데이터 분야의 초기 분석 단계로써 원 데이터를 분석하여 BA를 위한 사전 정보를 획득할 수 있으며, 이를 기반으로 비전문가도 BA의 다양한 툴 중에서 적합한 툴을 선택할 수 있는 정보를 제공하게 된다. 더불어 이러한 정보는 Info-Graphic의 다양한 시각화 툴 중에서 합리적이며 적절한 시각화 툴을 선택할 수 있는 정보를 제공하게 된다. 또한, 시

각화 기술은 공공 정보 LOD 등의 수집된 빅데이터를 하둡, 통계, 기계학습, 데이터마이닝 등의 처리 결과를 다양한 그래프 또는 차트로 이해하기 쉽게 표현하게 된다. InfograaS는 다양한 BA와 시각화 툴, 그리고 빅데이터가 필요한 다양한 새로운 영역에서도 사용될 것으로 예측한다.

### 감사의 글

본 논문은 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임 (2012R1A1A2041274)".

- [1] Owl Bus of Seoul Metro-city [Internet]. Available: <http://topis.seoul.go.kr/>.
- [2] P. Russom, "Big Data Analytics," TDWI Research, Fourth Quarter 2011 [Internet]. Available: <http://tdwi.org/blogs/philip-russom/2012/01/big-data-analytics-2012-new-years-predictions.aspx/>.
- [3] Cloud Spanning [Internet]. Available: <http://www.techopedia.com/definition/26534/cloud-spanning/>.
- [4] Public Data Portal [Internet]. Available: <https://www.data.go.kr/>.
- [5] D3 (Data-Driven Documentation) [Internet]. Available: <http://d3js.org/>.

### 참고문헌



#### 차 병 래 (Byung-Rae Cha)

2004년 2월 : 국립 목포대학교 컴퓨터 공학과(공학박사)  
 2005년 3월 ~ 2009년 2월 : 호남대학교 컴퓨터공학과 전임강사  
 2009년 9월~현재 : 광주과학기술원(GIST), 정보통신공학부 연구조교수  
 2012년 5월~현재 : 제노테크(주) 대표이사  
 관심분야 : 정보보안, Intrusion Detection System, 신경망, 클라우드 컴퓨팅, Future Internet 등



#### 이 형 호 (Hyung-Ho Lee)

1999년 2월: 호서대학교 생명공학과(이학사)  
 1999년 12월: 광운대학교 벤처창업대학원 수료  
 2000년 2월: LG멀티미디어스쿨 웹PD 과정 수료  
 2004년 2월 ~ 현재: (주)가민정보시스템 전략사업본부 본부장  
 관심분야: 웹, 클라우드 컴퓨팅, 유비쿼터스 컴퓨팅, 패턴인식 등



#### 심 수 정 (Su-Jeong Sim)

1996년 2월: 호남대학교 컴퓨터공학과 (공학사)  
 1999년 2월: 전남대학교 전산학과 소프트웨어공학 (공학석사)  
 2003년 8월: 전남대학교 전산학과 박사과정 수료  
 2012년 5월 ~ 현재: (주)동하테크 신입연구원  
 관심분야: 자연어 처리, 문서분류, 정보검색 등



#### 김 종 원 (Jong-Won Kim)

1997년 8월 ~ 2001년 7월 : University of Southern California 연구 조교수  
 1999년 12월 ~ 2000년 7월 Technology Consultant for VProtect Systems Inc.  
 2000년 7월 ~ 2001년 6월 Technology Consultant for Southern California Division of InterVideo Inc.  
 2001년 9월 ~ 2008년 3월 광주과학기술원 정보기전공학부 부교수  
 2008년 4월 ~ 현재 광주과학기술원 정보기전공학부 교수  
 관심분야 : Networked Media Systems and Protocols focusing "Reliable and Flexible Delivery for Integrated Media over Wired/Wireless Networks"