

빅데이터를 활용한 대한지적공사
사업효율 증대 및 이미지 제고 방안
(제36회 지적세미나)

2013. 11.

소 속 : 인천광역시본부 북부지사

직 급 : 기술직 6급

성 명 : 이 정 호

빅데이터를 활용한 대한지적공사
사업효율 증대 및 이미지 제고 방안

이 정 호

목 차

요 약	1
1. 서 론	3
1. 1. 연구의 배경	3
1. 2. 연구의 목적	3
1. 3. 연구의 범위	4
1. 4. 연구의 방법	4
2. 빅데이터의 이해	5
2. 1. 빅데이터의 출현 배경	5
2. 2. 빅데이터의 정의	6
2. 3. 빅데이터의 구성요소	8
가. 규모(Volume)	9
나. 속도(Velocity)	9
다. 다양성(Variety)	9
라. 복잡성(Complexity)	9
2. 4. 빅데이터의 특징	10
3. 빅데이터 분석 기법 및 기술	11
3. 1. 빅데이터 분석 기법	11
가. 데이터 마이닝(Data Mining)	12
나. 텍스트 마이닝(Text Mining)	13
다. 웹 마이닝(Web Mining)	13
라. 소셜 분석, 소셜 마이닝(Social Mining)	13
마. 현실 마이닝(Reality Mining)	14
바. 오피니언 마이닝(Opinion Mining)	14

3. 2. 빅데이터 분석 기술	15
가. 하둡(Hadoop)	15
나. R	16
다. NoSQL	17
3. 3. 소셜메트릭스를 활용한 대한지적공사 이미지 분석	19
4. 대한지적공사의 업무처리 현안	22
4. 1. 목표량 산정	22
4. 2. 민원처리	22
4. 3. 새업무 수주	23
5. 빅데이터의 활용에 따른 기대효과 및 과제	23
5. 1. 빅데이터의 활용 방안 및 기대효과	23
가. 목표량의 예측에 근거한 사업효율 증대 방안	23
나. 민원의 최소화를 통한 이미지 제고	25
다. 빅데이터를 활용한 수익 증대 방안	26
5. 2. 빅데이터 활용 주요 사례	28
가. 샌프란시스코, 범죄 예방 시스템으로 안전 지역사회 구축 사례	28
나. 유유제약의 ‘베노플러스’ 성공 사례	29
다. 공공기관 빅데이터 활용 사례	30
5. 3. 빅데이터 활용 과제	31
가. 전문 인력 양성	31
나. 개인정보 익명성 보장체계 확립	31
6. 결 론	32

표 목 차

[표 2.1] 빅데이터 정의	7
[표 2.2] 빅데이터 종류	8
[표 2.3] 빅데이터의 4가지 구성요소	10
[표 2.4] 빅데이터의 특성과 효과	10
[표 2.5] 미래사회의 특성과 빅데이터의 역할	11
[표 3.1] 빅데이터 분석(과거 vs 빅데이터 시대)	12
[표 3.2] 텍스트 마이닝 과정	13
[표 3.3] NoSQL 데이터 저장 시스템의 특징	17
[표 3.4] 빅데이터와 연계된 기술들	18
[표 5.1] 공공기관 빅데이터 활용 사례	30

그림 목 차

[그림 1.1] 기업의 빅데이터 분석 목적	4
[그림 2.1] 전세계 정보량 증가 추이	5
[그림 2.2] 정형-비정형 데이터 유형의 변화	8
[그림 3.1] 하둡의 구조와 분산처리 기술	16
[그림 3.2] 소셜메트릭스 소셜 검색 순위	19
[그림 3.3] 소셜메트릭스 소셜 인사이트 탐색어 맵	20
[그림 3.4] 소셜메트릭스 소셜 인사이트 탐색어 추이	20
[그림 3.5] 소셜메트릭스 소셜 인사이트 탐색어 여론	21
[그림 5.1] 목표량 산정 프로그램 가상화면	24
[그림 5.2] 도로공사 고객의 소리 분석시스템	25
[그림 5.3] 입찰정보시스템 및 LX IDEA TANK 가상화면	27
[그림 5.4] 샌프란시스코 경찰청의 범죄 지도(Crime Map)	28

빅데이터를 활용한 대한지적공사 사업효율 증대 및 이미지 제고 방안

이 정 호

요 약

최근 대한지적공사는 정보화 시대의 흐름에 발맞춰 많은 변화를 이뤄왔다. 자금 관리 시스템을 도입하기 이전, 금융업무의 집계는 서무의 보고를 통하지 않으면 본사에서 파악조차 어려웠으며 내부 ERP 시스템이 타시스템과 연계되지 않아 업무의 생산성이 저하되고, 금융자산 활용도 비효율적으로 운영되었다. 또한 도해지역의 측량성과와 측정점위치설명도를 확인하기 위해 지적서고를 쉼 세 없이 드나들었으며 접수자들은 수수료의 입금 여부를 확인하느라 은행사이트에서 손을 놓지 못하였다. 불과 5년도 지나지 않은 이야기다. 그러나 지금의 대한지적공사는 과거에 비해 혁신적인 업무시스템을 구축했으며 COS, MOS, ERP, CRM 등이 그 주인공이다. 모든 금융의 흐름은 본사와 본부에서 실시간으로 확인이 가능하여 자산의 효율적 운영이 가능해졌고, DRM 기술이 적용된 직원의 PC로 전산/전자화 된 결과도의 탐색이 가능하다. 또한 거래은행의 브랜치 솔루션이 적용된 COS에서 가상계좌로 입금된 상담업무의 즉각적인 접수가 가능하고, ERP는 전사적인 각종 자원의 신청, 관리를 총괄한다. 이는 사용자입장에서 실로 눈부신 발전이 아닐 수 없으며 우리의 업무처리를 보다 윤택하게 만들었다.

그러나 안타깝게도 이러한 전산자원의 발전이 아직까지는 우리의 정량적 수익을 증대시켜 주지는 못하고 있다. 자원의 축적과 사용은 가능하지만 이것으로 고부가가치의 새로운 성과 창출 영역을 만들어 내지는 못하고 있는 것이다. 이에 필요한 솔루션이 바로 빅데이터의 활용이다.

빅데이터란 데이터의 양, 생성 주기, 형식 등에서 과거 데이터에 비해 규모가 크고, 형태가 다양하여 기존의 방법으로는 수집·저장·검색·분석이 어려운 방대한 크기의 데이터를 말한다. 기업의 입장에서 빅데이터의 활용은 고객의 행동을 미리 예측하고 대처 방안을 마련하여 기업 경쟁력을 강화시키고, 생산성 향상과 비즈니스 혁신을 가능하게 하는 것으로 적용해보자면 우리는 잘 정제된 대한지적공사 내부의 축적자

원 뿐 아니라 무수히 많은 외부의 정형/비정형 데이터를 분석하여 가치 있는 데이터로 발전시킬 필요가 있는 것이다. IT활용을 전자화·자동화의 단계에서 멈춘 사업자와, 사업에 도움이 되는 인사이트 도출의 단계로 발을 내디딘 사업자는 미래사회에서 큰 경쟁력의 차이를 보일 것이다.

본 연구에서는 빅데이터의 출현 배경, 정의, 구성요소, 특징 및 분석 기술을 소개하고 대한지적공사의 업무처리 현안을 살펴보고 마지막으로 대한지적공사에서의 활용방안에 따른 기대효과와 선결 과제를 제시한다.

※ 인천광역시본부 북부지사

1. 서 론

1. 1. 연구의 배경

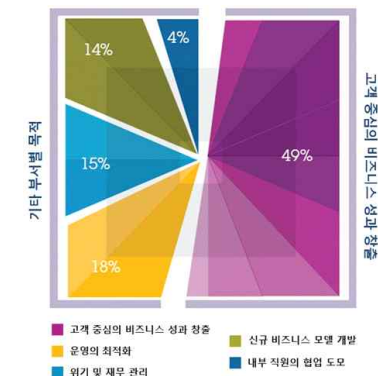
인간의 행동을 예측할 수 있는 새로운 수단으로 일컬어지는 빅데이터에 대한 관심이 높아지고 있다. 빅데이터는 그 양과 생성 주기, 형식 등이 지나치게 방대해서 기존의 방법으로는 처리하거나 분석하기는 어렵지만, 성질이 흡사 원유(原油)와 같아 잘만 가공하면 다양한 분야에서 최적에 요구 조건에 맞게 활용되어질 수 있는 차세대 기술을 일컫는다. 그리고 기업의 입장에서 빅데이터의 활용은 고객의 행동을 미리 예측하고 대처 방안을 마련하여 기업 경쟁력을 강화시키고, 생산성 향상과 비즈니스 혁신을 가능하게 한다. 이러한 빅데이터의 활용은 직관적 경험에 의한 업무처리가 주를 이루고 있는 우리 대한지적공사의 선진화 방안에 필수적 요소이며, 이를 위해서는 분석에 근거한 정량적 데이터를 기반으로 Total Solution을 도출해 내야한다. 2010년 MIT 슬로언 경영대학원에서 108개 국가 30개 업종의 경영진 3000명을 대상으로 설문조사를 한 결과에 따르면 고(高) 성과 기업들은 저(低) 성과 기업들과 비교할 때 직관보다 분석에 근거하는 비율이 평균적으로 5배 이상 높다는 보고가 그 이유이기도 하다. 또한 구글, 아마존, 페이스북, 애플 '4강'의 플랫폼 자체생산사업자 성공사례 뿐 만 아니라 국내 유수의 일반기업, 공공기관 성공사례가 주를 이루면서 빅데이터 도입의 필요성에 불을 붙이고 있는 것도 사실이다. 이에 대한지적공사는 빅데이터를 활용하여 얻을 수 있는 다양한 유형·무형의 수익 창출방안을 연구하여야 하며 지금이 공간정보 산업의 진흥을 선도할 글로벌 리더로서 확고히 자리매김할 수 있는 중요한 포지셔닝 기회다.

1. 2. 연구의 목적

빅데이터 환경이 보편화되면서 IT업계는 물론 일반기업, 공공기관의 빅데이터 경영 활동 역시 늘어나고 있다. 고객중심의 비즈니스 성과창출, 운영의 최적화, 위기관리 및 해결 등 다양한 경영활동 강화를 주요 분석 목적으로 삼고 있으며 데이터의 관찰 및 분석을 통해 새로운 시장을 발굴하고 차별화된 전략 수립을 기대하는 것이 그 이유다. 현재 갈수록 심해지는 지적측량 개방화 요구와 수치지역 및 지적확정측량에서의 업무 누수 등 대한지적공사의 수입과 입지는 과거와 비교하여 정비례하게 감소하

고 있으며 이에 대한 해결책 없이는 미래가 밝지 않은 것만은 확실하다. 이에 본 연구에서는 세계경제포럼(World Economic Forum), 2012년 국제 개발의 가장 큰 가능성을 여는 기술로 선정된 빅데이터의 정의 및 분석기술 등을 소개하며 이를 활용한 대한지적공사의 성과 창출 방안과 활용을 위한 선결 과제에 대해 제시하고자 한다.

[그림 1.1] 기업의 빅데이터 분석 목적



1. 3. 연구의 범위

본 연구의 범위는 현재 사용되고 있는 빅데이터 분석 기술 정리와 분석 방법에 근거한 대한지적공사의 빅데이터 적용 가능 영역이다. 급속성장의 중심에 서있는 IT기술의 특성상 미래에 나타날 빅데이터 신기술의 범위와 생성 주기를 예측하는 것은 상당한 어려움이 있다. 그러나 이는 반대로 대한지적공사에서 활용 가능한 획기적인 성과창출방안 역시 무궁무진하다는 것을 뜻한다.

1. 4. 연구의 방법

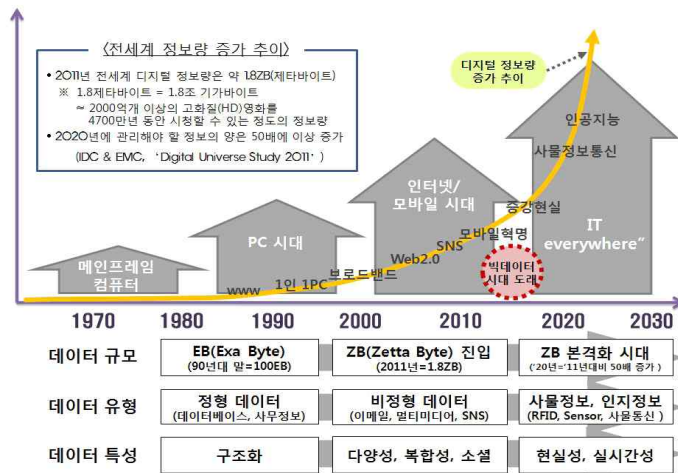
본 연구의 방법은 도서 및 논문, 보고서를 통한 빅데이터의 이론 고찰과 분석기술을 적용한 성공사례분석을 기준으로 작성하였으며 분석기술을 통한 성과 창출 방안에 주안(主眼)을 두어 연구를 진행하였다. 또한 민원해결 및 이미지 제고에 획기적인 역할을 담당할 소셜 마이닝(Social Mining) 기술을 활용해 보고자 다음소프트의 소셜 메트릭스를 이용해 대한지적공사의 이미지 분석을 시도하였다.

2. 빅데이터의 이해

2. 1. 빅데이터의 출현 배경

전 세계적으로 생산, 유통되는 디지털 정보량이 2011년 1.8 제타바이트(ZB)¹⁾, 약 2조 기가바이트에 달하고 있다. 이는 전 세계 인구 1인당 평균적으로 280기가바이트, 약 45편의 DVD 영화(1편당 6기가바이트 기준)에 해당하는 엄청난 정보량인 셈이다. 빅데이터 현상은 기업들의 고객 데이터 수집활동 및 멀티미디어 콘텐츠의 폭발적 증가와 스마트폰 보급으로 인한 SNS 활성화 및 사물통신망의 저변 확대로 빠르게 확산되고 있다. 데이터 처리는 기존에도 논의되어 왔지만 최근 다시 주된 관심사로 떠오르는 것은 IT를 활용한 다양한 산업분야의 전자화나 자동화가 크게 진전됨으로써 처리 및 분석에 활용할 수 있는 대량의 데이터가 축적되고 있기 때문이다.

[그림 2.1] 전세계 정보량 증가 추이



1) 바이트 크기

기호	KB	MB	GB	TB	PB	EB	ZB	YB
이름	킬로바이트	메가바이트	기가바이트	테라바이트	페타바이트	엑사바이트	제타바이트	요타바이트
값	1024B	1024KB	1024MB	1024GB	1024TB	1024PB	1024EB	1024ZB

세계적인 컨설팅 기관인 Mckinsey and Company는 'Big Data : The Next Frontier for Innovation, Competition, and Productivity'에서 빅데이터 부상배경을 아래와 같이 정리하고 있다.

첫째, 기업의 고객 데이터 트래킹 및 수집행위의 증가이다. 고객 데이터들이 인터넷, 스마트폰 등 다양한 매체들을 통해 데이터가 트래킹 되고, 온라인뿐만 아니라 오프라인상에서도 사용자 정보, 소비자 행태에 대한 정보 수집이 가능해 졌다.

둘째, 저장매체와 카메라 모듈, 디스플레이 가격의 인하는 멀티미디어 콘텐츠 사용 확산과 이에 관한 정보의 증가를 가져왔다. 고화질 동영상은 이미 인터넷 전체 트래픽의 50% 이상을 차지하며 2013년 70%로 증가할 전망이다.

셋째, 트위터, 페이스북 등 SNS의 급격한 확산과 더불어 텍스트 등 비정형 데이터의 폭증이다. 페이스북에서만 매일 한 이용자당 평균 90개 이상의 콘텐츠가 업로드가 되고 있으며, YouTube에서는 1분마다 24시간 분량의 비디오가 업로드 된다. SNS에서 유통되는 정보는 비정형 데이터로서 이의 처리를 위해서는 추가적인 데이터처리가 필요하여 데이터의 복잡성을 증가시킨다.

넷째, M2M(Machine to Machine) 및 IoT(Internet of Things) 등의 통신 기술 발전에 따른 사물 간 통신 네트워크에서 발생하는 데이터의 증가이다. M2M, IoT 등의 활성화를 통해 데이터를 사용자가 생성하지 않고, 인프라 자체가 대량의 데이터를 생성하게 되었다.

2. 2. 빅데이터의 정의

빅데이터는 기존의 데이터 수집·저장·관리·분석 역량을 넘어서는 대량의 데이터 세트를 의미하며, 기존의 관계형 데이터와 비교하여 양, 속도, 다양성 및 복잡성에 그 차이를 볼 수 있다. 데이터에는 정형화된 데이터와 비정형화된 데이터가 있는데, 최근에 논의되고 있는 빅데이터는 정형화된 것이든 아니든 상관없이 엄청난 양의 데이터를 말한다. 빅데이터에 대한 정의는 다양하지만, 기업적인 측면에서 빅데이터는 기업의 효과적인 전략 도출에 필요한 상세하고 높은 빈도로 생성되는 다양한 종류의 데이터로 정의할 수도 있다. 다음은 각 기관에서 내린 빅데이터의 정의이다.

[표 2.1] 빅데이터 정의

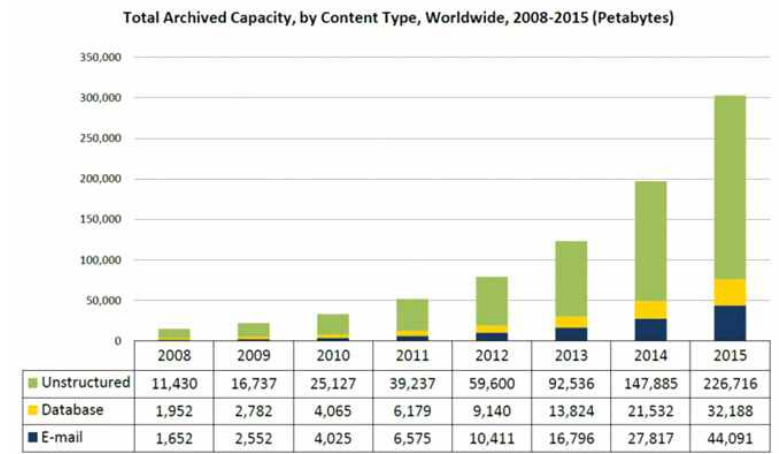
기관	빅데이터 정의	시사점
McKinsey (2011)	일반적인 데이터베이스 소프트웨어가 저장, 관리, 분석할 수 있는 범위를 초과하는 규모의 데이터	데이터 규모에 초점(정량적 측면 강조)
IDC (2011)	다양한 종류의 대규모 데이터로부터 저렴한 비용으로 가치를 추출하고 데이터의 초고속 수집, 발굴, 분석을 지원하도록 고안된 차세대 기술 및 아키텍처	데이터 규모가 아닌 업무 수행에 초점 3V(Volume, Velocity, Variety) 또는 4V(3V+Value)
Wikipedia	기존 데이터베이스 관리도구의 데이터 수집·저장·관리·분석의 역량을 넘어서는 대량의 정형 또는 비정형 데이터 세트 및 이러한 데이터로부터 가치를 추출하고 결과를 분석하는 기술	데이터 규모 및 업무 수행의 관점에서 통합된 정의
Gartner	기존 3V(Volume, Velocity, Variety)에 복잡성(Complexity) 추가	구조화되지 않은 데이터, 데이터 저장 방식 차이, 중복성 문제 등 데이터 관리 및 처리 복잡성 심화로 복잡성을 특성으로 추가
IBM	기존 3V(Volume, Velocity, Variety)에 데이터의 진실성(Veracity) 추가	진실성이 확보된 데이터를 바탕으로 분석해야 한다는 데이터 품질 측면의 특성 추가
최근 동향	수많은 정형 데이터 혹은 비정형 데이터를 수집하면, 분명한 패턴이 나오게 되며, 이를 통해 수집된 데이터를 기반으로 한 예측 분석	매출 증가, 비용 절감, 고객만족 증대라는 비즈니스 가치를 창출할 수 있는 패턴 발견에 집중

빅데이터에 대한 정의는 보는 관점에 따라 다양하지만, 결론적으로 모두가 말하고자 하는 것은 빅데이터란 현재 시스템의 처리 범위를 넘어서는 데이터이며, 기존의 방식과 다른 새로운 처리 및 분석방법이 필요하다는 것이다. 그리고 빅데이터는 [표 2.2]와 같이 데이터의 정형화 정도에 따라 정형(Structured), 반정형(Semi-Structured), 비정형(Unstructured) 데이터로 분류된다. 이중 기업 내부에 축적된 비정형 데이터의 비중이 2006년 36%에서 2015년에는 77%로 2배 이상 급증할 것으로 전망하고 있으며, 비교적 선형적으로 증가하던 정형 데이터조차 연간 40~60%에 이르는 증가세를 보이고 있다.

[표 2.2] 빅데이터 종류

정의	설 명
정형 (Structured)	◦ 고정된 필드에 저장된 데이터 (ex) 관계형 데이터베이스, 스프레드시트
반정형 (Semi-Structured)	◦ 고정된 필드에 저장되어 있지는 않지만, 메타데이터나 스키마 등을 포함하는 데이터 (ex) XML이나 HTML 텍스트
비정형 (Unstructured)	◦ 고정된 필드에 저장되어 있지 않은 데이터 (ex) 텍스트 분석이 가능한 텍스트 문서 및 이미지/동영상/음성 데이터

[그림 2.2] 정형-비정형 데이터 유형의 변화



Source: Enterprise Strategy Group, 2010.

2. 3. 빅데이터의 구성요소

빅데이터의 구성요소는 일반적으로 3V(Volume, Velocity, Variety)를 기본으로 1V(Value)나 1C(Complexity)가 추가되어 설명되나 구조화되지 않은 데이터, 데이터의 저장방식 차이, 중복성 문제 등 데이터의 관리 및 처리 복잡성 심화로 1C(Complexity)를 특성으로 추가한 Gartner의 기본 정의에 의한 구성요소를 표준으로 보고 있다.

가. 규모(Volume)

데이터의 발생량으로 페타바이트, 제타바이트 이상의 정보를 기준으로 물리적인 크기 뿐 만 아니라 데이터의 속성에도 연관되어 있으며 그것을 처리하는데 어려움이 있는지 혹은 없는지를 의미한다. 오늘날 다양한 문제 해결에 필요한 훨씬 복잡 미묘한 패턴을 잡아내기 위해 규모의 중요성은 더욱 커지고 있다.

나. 속도(Velocity)

데이터를 처리하는 속도로 배치 분석만을 의미하는 것이 아니라 필요에 따라서 수많은 사용자 요청을 실시간으로 처리한 후 처리 결과를 반환해주는 기능에 대한 시간적 의미와 연관된다. 다시 말해서, 데이터의 생성과 처리가 진행되는 속도를 말하며, 의미는 세 가지로 구분된다. 데이터가 발생 후 기업 내의 스토리지에 저장되기까지의 속도와 발생한 데이터의 불필요 부분과 무의미한 부분을 처리하여 가용하게 되는 수준까지의 속도, 그리고 정제된 데이터를 분석하고 의미를 추출하여 최종 목적을 달성하는 속도까지 말한다. 이러한 속도는 데이터의 접근성과 사용 가능성을 높이는 데 많은 영향을 준다.

다. 다양성(Variety)

기존의 전통적인 기업분석은 데이터베이스에서 관리하는 구조적인 데이터로 수치 및 텍스트가 주를 이루었다. 하지만 빅데이터 분석에는 통일된 구조로 정리하기 어려운 비정형 데이터 분석이 전체 중 90% 이상을 차지하며 사진, 동영상 등 기존의 구조화 된 데이터가 아닌 다양한 형태의 데이터가 포함된다. 빅데이터를 수집하고 분석할 수 있게 되면서 훨씬 다양한 데이터의 활용이 가능해졌으며 특히, 다양한 비정형 데이터에서 생각지도 못했던 결과를 얻는 일이 빈번해졌다.

라. 복잡성(Complexity)

데이터의 복잡성이란 데이터 구조, 도메인 규칙, 저장 타입 등 데이터의 발생, 처리, 정제 등의 과정에 포함된 모든 요소가 복잡해지는 것을 의미한다. 예를 들어 미디어 관리 시스템에서 영상 데이터는 다양한 인코딩 포맷을 하고 있기 때문에 메타 데이터를 기준으로 데이터를 분류하고 파싱(Parsing)하는 작업이 필요하다. 이처럼 데이터 유형에 따라 분석을 위한 전처리(Preprocessing)가 필요할 수 있다.

[표 2.3] 빅데이터의 4가지 구성요소

구분	주요 내용
규모의 증가 (Volume)	◦ 기술적인 발전과 IT의 일상화가 진행되면서 해마다 디지털 정보량이 기하급수적으로 폭증 ⇒ 제타바이트(ZB) 시대로 진입
속도의 증가 (Velocity)	◦ 사물정보(센서, 모니터링), 스트리밍 정보 등 실시간성 정보 증가 ◦ 실시간성으로 인한 데이터 생성, 이동(유통) 속도의 증가 ◦ 대규모 데이터 처리 및 가치 있는 현재정보(실시간) 활용을 위해 데이터 처리 및 분석 속도가 중요
다양성의 증가 (Variety)	◦ 로그기록, 소셜, 위치, 소비, 현실데이터 등 데이터의 종류의 증가 ◦ 텍스트 이외의 멀티미디어 등 비정형화된 데이터 유형의 다양화
복잡성의 증가 (Complexity)	◦ 구조화되지 않은 데이터, 데이터 저장방식의 차이, 중복성문제 등 ◦ 데이터 종류의 확대, 외부 데이터의 활용으로 관리대상의 증가 ◦ 데이터 관리 및 처리의 복잡성이 심화되고 새로운 기법 요구

2. 4. 빅데이터의 특징

빅데이터는 데이터의 엄청난 크기, 빠른 전송, 다양성 및 비정형성, 복잡성과 같은 특징을 가지는데, 이는 기존의 기술과 방법으로는 ‘다루기 어려움’으로 요약이 가능하다. 이와 같은 어려움에도 불구하고, 빅데이터는 기업과 정부에 새로운 발전의 기회를 제공하거나 전 지구적 문제를 해결할 수 있는 단초를 제공할 것으로 기대되는, ‘새로운 지적 통찰력의 기회’로 인식되고 있다.

[표 2.4] 빅데이터의 특성과 효과

특성	효과
대규모 (Huge Scale)	◦ 기술 발전으로 데이터를 수집, 저장, 처리 능력 향상 ◦ 현실세계 데이터를 기반으로한 정교한 패턴분석 가능 ◦ 데이터가 많을수록 유용한 데이터, 전혀 새로운 패턴의 정보를 찾아낼 수 있는 확률도 증가
현실성 (Reality)	◦ 우리사회 일상에서의 데이터 기록물의 증가 등 현실 정보, 실시간 정보의 축적이 급증될 전망 ◦ 개인의 경험, 인식, 선호 등 인지적인 정보 유통 증가
시계열성 (Trend)	◦ 현시점 뿐 만 아니라 과거 데이터의 유지로 시계열적인 연속성을 갖는 데이터의 구성 ◦ 과거, 현재, 미래 등 시간 흐름상의 추세 분석 가능
결합성 (Combination)	◦ 의료, 범죄, 환경, 안보 등 타분야, 이종 데이터간의 결합으로 새로운 의미의 정보 발견 ◦ 실제 물리적인 결합 이전에, 데이터의 결합을 통한 사전 시뮬레이션, 안전성 검증 분야 발전 가능

빅데이터를 잘 활용하면 미래사회에서 새로운 기회를 창출하고, 위험을 해결하는 사회 발전의 엔진 역할을 수행할 것으로 기대된다. 사회 발전의 속도가 빨라지고, 위험요인과 복잡성이 증가할수록 시스템적으로 신속하게 환경 변화를 감지할 필요성이 증가하며 빅데이터의 분석을 통해 미래의 통찰력, 대응력, 경쟁력, 창조력을 향상시키며 국가의 지속발전을 이룰 수 있는 전략 수립이 필요하다.

[표 2.5] 미래사회의 특성과 빅데이터의 역할

미래사회의 특성	빅데이터의 역할	
불확실성	⇒ 통찰력	- 사회현상, 현실세계의 데이터를 기반으로 한 패턴분석과 미래전망 - 여러 가지 가능성에 대한 시나리오 시뮬레이션 - 다각적인 상황이 고려된 통찰력을 제시 - 다수의 시나리오로 상황 변화에 유연하게 대처
리스크	⇒ 대응력	- 환경, 소셜, 모니터링 정보의 패턴 분석을 통한 위험 징후, 이상 신호 포착 - 이슈를 사전에 인지·분석하고, 빠른 의사결정과 실시간 대응 지원 - 기업과 국가 경영의 투명성 제고 및 낭비요소 절감
스마트	⇒ 경쟁력	- 대규모 데이터 분석을 통한 상황인지, 인공지능 서비스 등 가능 - 개인화, 지능화 서비스 제공 확대 - 소셜(니즈)분석, 평가, 신용, 평판 분석을 통해 최적의 선택 지원 - 트렌드 변화 분석을 통한 제품 경쟁력 확보
융 합	⇒ 창조력	- 타분야와의 결합을 통한 새로운 가치창출 - 인과관계, 상관관계가 복잡한 컨버진스 분야의 데이터 분석으로 안전성 향상, 시행착오 최소화 - 방대한 데이터 활용을 통한 새로운 융합시장 창출

3. 빅데이터의 분석 기법 및 기술

3. 1. 빅데이터의 분석 기법

빅데이터 시대에는 데이터가 IT에서 분리된 독립적인 주체로 발전하며 IT의 주도권이 인프라, 기술, S/W에서 데이터로 전이되었다. 최근 빅데이터가 이슈가 되자 데이터 분석의 중요성에 대한 인식과 다양한 데이터 마이닝 기법이 새롭게 조명되고 있다.

[표 3.1] 빅데이터 분석(과거 vs 빅데이터 시대)

분석속성	과거	빅데이터 현재
데이터세트	사전정의	포괄적, 반복적
데이터속도	배치	실시간
데이터분석	과거분석	예측, 시계열, 최적화
인프라형태	독립인프라	리소스 풀
아키텍처	최적화	분산/병렬/인메모리 연계
데이터크기	수십 GB	수십, 수백 TB 이상
데이터종류	정형데이터	반정형, 비정형데이터 포함
복잡도	특정주제	예측 및 최적화 관점의 요건
분석의 양	수십개의 모델	수천, 수만개 이상의 세분화 모델
분석사용자	분석 전문가	데이터 시각화, 분석 모델링 자동화

마이닝(mining)이란 광산에서 광물을 캐낸다는 의미로써, 정보화 사회인 현재는 다량의 적재된 데이터에서 숨겨진 패턴과 관계 등을 의사결정이나 미래를 전망할 수 있는 유용한 정보로 추출한다는 의미로 사용되고 있으며, 기업의 의사결정, 마케팅, 고객 관리 뿐 만 아니라 금융, 의학, 교육, 환경 등의 분야에서 혁신적으로 적용되고 있다.

가. 데이터 마이닝(Data Mining)

데이터 마이닝은 대용량의 데이터, 데이터베이스 등에서 감춰진 지식, 기대하지 못했던 경향, 새로운 규칙 등의 유용한 정보를 발견하는 과정으로 데이터 마이닝을 통해 정보의 연관성(순차 패턴, 유사성 등)을 파악함으로써 가치 있는 정보를 만들어 의사결정에 적용하는 기술이다. 다시 말해 기업이 보유하고 있는 일일 거래자료, 고객자료, 상품자료, 마케팅 활동의 피드백 자료와 기타 외부자료를 포함하여 사용 가능한 데이터를 기반으로 숨겨진 지식, 기대하지 못했던 패턴, 새로운 법칙과 관계를 발견하고 이를 실제 경영의 의사결정 등을 위한 정보로 활용하자 하는 것이다.

현재 대한지적공사에서 운영하고 있는 COS, MOS, CRM, ERP, CDP등에 축적된 데이터를 정량적으로 사용하는 것 뿐 아니라 이를 이용해 미래의 가치를 창출하고 합리적 의사결정을 하는데 반영하는 것 등의 기술이 데이터 마이닝이다.

나. 텍스트 마이닝(Text Mining)

텍스트 마이닝은 비/반정형 텍스트 데이터에서 자연어처리(Natural Language Processing)기술을 기반으로 유용한 정보를 추출, 가공하는 것을 목적으로 하는 기술이다. 텍스트 마이닝 기술을 통해 방대한 텍스트 문치에서 의미 있는 정보를 추출해 내고, 다른 정보와의 연계성을 파악하며, 텍스트가 가진 카테고리를 찾아내거나 단순한 정보 검색 그 이상의 결과를 얻어낼 수 있다. 컴퓨터가 인간이 사용하는 언어(자연어)를 분석하고 그 안에 숨겨진 정보를 발굴해 내기 위한 대용량 언어자원과 통계적, 규칙적 알고리즘이 사용되고 있다.

[표 3.2] 텍스트 마이닝 과정

과정	설 명
Text 문서	◦ DB Contents, 텍스트 기반의 문서(웹 문서, 오피스문서, 메일 등)
Text 전처리	◦ 문서 내에 표현되어 있는 단어·구·절에 해당하는 내용을 언어분석·처리 과정으로 가공하여 데이터로 표현
의미정보변환	◦ 전처리된 데이터 중 의미 있는 정보를 선별하여 저장(불용어처리, 대소문자 처리, Stemming처리)
의미정보추출	◦ 복잡한 의미정보의 표현을 단순화하고 도메인에 적합한 정보를 문서의 의미 데이터로 저장
패턴 및 경향분석	◦ Feature정보를 기반으로 문서를 자동 군집하거나 자동 분류하는 등의 정보 재생산
정보표현 및 평가	◦ 새롭게 생성된 정보를 사용자에게 시각화하여 효과적으로 표현, 평가과정을 통해 텍스트 마이닝의 처리과정 중 문제가 되는 부분을 수정 및 보완하여 품질 및 성능을 높이는 데 활용

다. 웹 마이닝(Web Mining)

인터넷상에서 수집된 정보를 데이터 마이닝 방법으로 분석통합하는 기법으로 웹 마이닝은 콘텐츠 마이닝(웹 검색, 수집 데이터), 구조 마이닝(웹 사이트 구조), 활용 마이닝(사용자 이용형태) 등으로 세분될 수 있다.

라. 소셜 분석, 소셜 마이닝(Social Mining)

소셜 네트워크의 분석은 수학의 그래프 이론에 뿌리를 두고 있으며, 소셜 네트워크 연결구조 및 연결 강도 등을 바탕으로 사용자의 명성 및 영향력을 측정하여, 소셜 네

트워크상에서 입소문의 중심이나 허브 역할을 하는 사용자를 찾는 데 주로 활용된다. 소셜 미디어에 올라오는 글과 사용자를 분석해 소비자의 흐름이나 패턴 등을 분석하고, 판매나 홍보에 적용하여 마케팅 분야 뿐 만 아니라 사회의 흐름과 트렌드, 여론 변화 추이를 읽어내는 소셜 미디어 시대의 새로운 마이닝 기법이다.

소셜 네트워크 분석 기술은 이전부터 있었으나 20세기 말 네트워크 사이언스의 출현으로 네트워크 분석 기술이 급격히 발전하게 되면서, 그 결과, 기존 기술로는 불가능했던 대규모 소셜 네트워크 데이터 분석을 빠르고 정확하게 할 수 있게 된 것이다.

SNS는 개인을 노드(Node), 개인의 사회적 관계를 링크(Link)로 간주하면 소셜 네트워크를 구축할 수 있고 이렇게 형성된 소셜 네트워크에서 다음의 4단계를 통해 정보를 추출 및 분석할 수 있다. 첫째, 소셜 네트워크의 위상학적 구조(Network Topology Structure)분석으로 네트워크 전반적 특성을 파악한다. 둘째, 네트워크 구조의 시간에 따른 진화를 분석한다. 셋째, 네트워크상의 각 노드(사용자)가 생산, 확산시키는 콘텐츠(포스트, 댓글, 리트윗, 동영상, 링크 등) 흐름을 분석한다. 넷째, 종합하여 각 개인 또는 그룹의 소셜 네트워크 내 영향력, 관심사, 성향 및 행동 패턴을 분석 추출한다. 소셜 네트워크 분석의 활용 효과는 이미 각 기업이 빅데이터에 주목하면서 SNS 데이터 분석기술을 통해 방대한 비정형 데이터들을 분석하고 이를 비즈니스에 활용하고 있다. 하지만 그 부작용 최소화를 위한 자체 모니터링, 위험 완화 프로그램 개발 등 관련 기업·시장의 사회적 책임이 강조되어야 하며, 프라이버시 보호 등 부작용대응을 위한 기술개발 및 산업육성 지원이 이루어져야 한다.

마. 현실 마이닝(Reality Mining)

사람들의 행동패턴을 예측하기 위해 사회적 행동과 관련된 정보를 기기(휴대폰, GPS 등)를 통해 얻고 분석하는 기술이다. 휴대폰 등 모바일 기기들을 통해 현실에서 발생하는 정보를 기반으로 인간관계와 행동 양태 등을 추론한다.

바. 오피니언 마이닝(Opinion Mining)

최근 새로운 여론 분석기술로 주목받고 있는 오피니언 마이닝은 소셜미디어와 웹 사이트 등에 나타난 여론과 의견을 분석하여 유용한 정보로 재가공하는 기술이다. 오피니언 마이닝 혹은 감정분석을 활용하면 네티즌이 그들에 대해 이야기하는 댓글이나 포스팅 등을 긍정, 부정, 중립으로 분류하여 더 객관적이고 정확하게 평판을 파악

할 수 있다. 오류를 피하기 위해서 사용되는 오피니언 마이닝 연구의 3단계는 다음과 같다. 첫째, ‘주관성 분석’으로 주어진 텍스트가 주관적인지 객관적인지 결정하는 것으로, 주어진 텍스트에 나타난 저자의 태도를 판단하는 단계이다. 둘째, ‘극성 분석’으로 텍스트가 주관적인 의견을 갖고 있을 경우 긍정적인 부정인지를 분류하는 단계이다. 셋째, ‘극성의 정도 분석’으로 주관적인 텍스트에 대하여 긍정적인 정도와 부정적인 정도를 측정하는 단계이다. 오피니언 마이닝은 특정 서비스 및 상품에 대한 시장 규모예측, 소비자의 반응, 입소문 분석 등에 활용되고 있으며, 공공분야의 경우 민원의 원인이나 문제점 등을 파악하는 것이 용이해 서비스를 개선할 수 있다. 또한, 기업은 특정 제품에 대한 고객의 반응을 빠르게 파악하고 선호도를 역으로 추론하는데 효과적으로 활용할 수 있다. 정확한 오피니언 마이닝을 위해서는 전문가에 의한 선호도를 나타내는 표한·단어의 축적이 필요하며 대한지적공사는 국가의 공간정보 업무를 선도하는 공공기업으로서 업무처리의 질을 높이는 것은 물론 민·관으로 부터의 긍정적 이미지를 이끌어내는 것이 민원을 줄이고 미래 사업개발을 해나가는 데 중요한 추진 전략이 될 것이다.

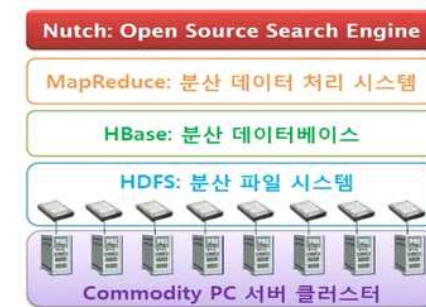
3. 2. 빅데이터의 분석 기술

가. 하둡(Hadoop)

하둡은 다수의 서버를 묶어 분산 처리하는 플랫폼이며, 분산처리를 하는 ‘Map’ 단계와 결과를 취합하는 ‘Reduce’ 단계로 이루어진 구글의 MapReduce 모델을 본떠 만든 것으로 Apache Group에 의해 제작되고 배포되었다. 구글 내부에서 사용하던 하둡이 오픈 소스로 발표되어 크게 주목을 받으면서 사실상 야후와 페이스북, 트위터 등 많은 기업의 빅데이터 처리를 위해서 사용되고 있다.

하둡의 주요 구성요소 [그림 3.1]와 같이 분산 파일 시스템인 HDFS, 분산처리를 위한 프레임워크 MapReduce가 핵심이다. HDFS는 페타바이트 이상의 데이터를 저장할 수 있는 분산 파일시스템으로 고가의 서버가 아닌 범용 x86 서버와 서버당 다수의 로컬디스크를 사용하여 파일 시스템을 구성한다. 구글에 의해 고안된 MapReduce는 함수형 프로그래밍 언어인 LISP를 모델로 하여 Map과 Reduce라는 2개의 함수를 이용하여 대용량 데이터를 쉽고 빠르게 처리할 수 있는 방법이다.

[그림 3.1] 하둡의 구조와 분산처리 기술



하둡 자체는 파일 시스템과 분산 처리 플랫폼이지만 하둡을 중심으로 다양한 에코 시스템이 구축되어 있다. 분산 데이터베이스인 Hbase, 검색엔진 너치(Nutch), 대규모 데이터셋 탐색용 데이터 흐름 언어와 실행 환경 및 관계형 대수 쿼리 언어 인터페이스인 피그(Pig), HDFS에 저장된 데이터를 관리하는 데이터웨어하우징 솔루션인 하이브(Hive) 등이 포함되어 있다. 하둡은 아파치 소프트웨어 라이선스에 따라 전체 프레임 워크가 오픈소스이기 때문에 기본 소프트웨어에 대한 라이선스 비용이 없고, 또한 고가의 하드웨어나 고성능 프로세서가 필요하지 않고 저가의 상용 서버를 사용할 수 있어 비용 절감 효과가 높은 것으로 평가되고 있다.

나. R

오픈소스 프로젝트 R은 통계 계산 및 시각화를 위한 언어 및 개발환경을 제공하며, R 언어와 개발환경을 통해 기본적인 통계 기법부터 모델링, 최신 데이터 마이닝 기법까지 구현/개선이 가능하다. 이렇게 구현한 결과는 그래프 등으로 시각화할 수 있으며, Java나 C, Python 등의 다른 프로그래밍 언어와 연결도 용이하다. 또 R의 경우에는 GNU 프로젝트로 만들어져서 무료로 사용될 수 있고, 다양하게 활용할 수 있는 패키지들이 빠르게 추가되고 있어 최근 들어 더욱 주목을 받고 있다. 위의 장점으로 인해 R은 통계분석 분야에서 인지도를 높여왔으며, 하둡 환경상에서 분산처리를 지원하는 라이브러리 덕분에 구글, 페이스북, 아마존 등의 빅데이터 분석이 필요한 기업에서 대용량 데이터 통계분석 및 데이터 마이닝을 위해 널리 사용되고 있다.

다. NoSQL

NoSQL은 Not-Only SQL의 줄임말로써 기존 관계형 특징(스키마, 테이블간의 조인 연산 등)을 포함하지 않고 다른 특징을 가지는 데이터베이스를 일컫는 말이다. 즉, 전통적인 관계형 데이터베이스와 다르게 설계된 비관계형 데이터 베이스를 의미한다. 이러한 NoSQL은 데이터의 양이 증가하고 저장할 데이터가 지속적으로 변화하며 데이터의 요구가 일관적이지 않은 웹 시장의 요구를 수용하고, 기존의 관계형 데이터베이스의 수평적 확장성의 한계를 해결하고자 대두되었다. 관계형 데이터베이스의 경우, 일관성(Consistency : 모든 노드는 같은 시간에 같은 데이터를 보여줘야 한다)과 유효성(Availability : 일부 노드가 다운되어도 다른 노드에 영향을 주지 말아야 한다)에 중점을 두고 있는 반면, NoSQL 기술은 분산가능성(Partition Tolerance : 네트워크 전송 중 일부 데이터를 손실하더라도 시스템은 정상 동작을 해야 한다)에 중점을 두고 일관성과 유효성은 보장하지 않는다. 이것은 일관성, 유효성, 분산가능성 중 2가지만 보장이 가능하다는 분산 데이터베이스 시스템 분야의 CAP 이론에 따른 것이다. 따라서 대규모의 유연한 데이터 처리를 위해서는 NoSQL 기술이 적합하지만, 안정성이 중요한 시스템에서는 오랫동안 검증된 관계형 데이터베이스를 채택할 필요가 있다.

[표 3.3] NoSQL 데이터 저장 시스템의 특징

NoSQL 데이터 저장 시스템의 특징
◦ Key & Value로 저장
◦ 분산 환경 지원
◦ Call level interface 지원(DBMS에 접근하는 표준)
◦ 막대한 양의 데이터를 처리할 수 있는 대용량 데이터의 빠른 인덱싱
◦ 클러스터나 그리드에서의 구동을 위해 다양한 테이블로 데이터베이스를 나눠야 하는 복잡한 작업과 '샤딩(sharding)' 없이 손쉽게 저렴하게 여러 서버들의 수평적 확장(horizontal scaling)이 가능
◦ 데이터의 스키마와 속성들을 동적으로 정의

[표 3.4] 빅데이터와 연계된 기술들

용어	설명
빅 테이블 (Big Table)	◦ 구글 파일 시스템 상에 구축된 상용 분산 데이터베이스 시스템 ◦ H베이스에 영향을 미침
카산드라 (Cassandra)	◦ 분산 시스템에서 방대한 분량의 데이터를 처리할 수 있도록 디자인 된 오픈소스 데이터베이스 관리시스템 ◦ 이 시스템은 원래 페이스북에서 개발했으며 지금은 아파치 소프트웨어 재단의 한 프로젝트로 관리되고 있음
데이터웨어하우스 및 분석 어플라이언스	◦ 데이터웨어하우스를 위해 서버, 스토리지, 운영체제, 데이터베이스, BI, 데이터마이닝 등 기타 여러 가지 소프트웨어가 최적화되어 설치된 통합제품
분산시스템	◦ 동시에 일을 처리하기 위해 네트워크로 연결된 컴퓨터들의 집합 ◦ 단일 또는 다수의 컴퓨터 리소스를 부분적으로 활용함으로써 시스템의 가격 대비 성능비, 안정성, 확장성을 향상시킬 수 있음
구글 파일 시스템	◦ 구글에서 개발한 분산파일 시스템으로 하둡(Hadoop)과 관련이 있음
하둡 (Hadoop)	◦ 분산 시스템 상에서 대용량 데이터 처리 분석을 지원하는 오픈소스 소프트웨어 프레임워크로 구글이 개발한 맵리듀스를 오픈소스로 구현한 결과물 ◦ 야후에서 최초 개발되었으며 지금은 아파치 소프트웨어 재단의 한 프로젝트로 관리되고 있음
H베이스 (Hbase)	◦ 구글의 '빅 테이블'을 참고로 개발된 오픈소스 분산 비관계형 데이터베이스 ◦ 파워셋에서 개발했으며 현재는 아파치 소프트웨어 재단에서 하둡(Hadoop)의 일관인 프로젝트로 관리되고 있음
맵리듀스 (MapReduce)	◦ 분산 시스템 상에서 대용량 데이터 세트를 처리하기 위해서 구글이 제안한 소프트웨어 프레임워크, 하둡(Hadoop)에도 구현 되었음
비관계형 데이터베이스	◦ 비관계형 데이터베이스는 데이터를 테이블에 저장하지 않은 데이터베이스이며 관계형 데이터베이스와는 대조적인 개념 ◦ 이를 사용하면 스키마 없는 엔터티(NoSQL)를 관리할 수 있음

3. 3. 소셜메트릭스를 활용한 대한지적공사 이미지 분석

소셜미디어는 기존의 일방적인 매스미디어를 보완하기 위해 나온 양방향적인 도구이다. 소셜미디어에서 우리 기업을 검색하고 모니터링해서 분석한 다음 우리가 잘 하고 있는 점은 더욱 강화하고, 고쳐야 할 점은 보완해서, 보다 좋은 서비스를 제공한다면 별다른 홍보나 마케팅 없이도 자연적인 홍보 효과 및 긍정적 이미지 효과를 얻을 것이다. 하지만 거대 소셜 미디어상의 데이터를 일일이 찾아서 분석한다는 것은 결코 만만한 일이 아니다. 그래서 이러한 모니터링과 분석을 수월하게 할 수 있도록 도움을 주는 도구인 다음소프트의 소셜메트릭스를 활용하여 2013년 4월 13일에서 2013년 5월 13일까지의 대한지적공사 여론 분석추이를 살펴보았다.

[그림 3.2] 소셜메트릭스 소셜 검색 순위



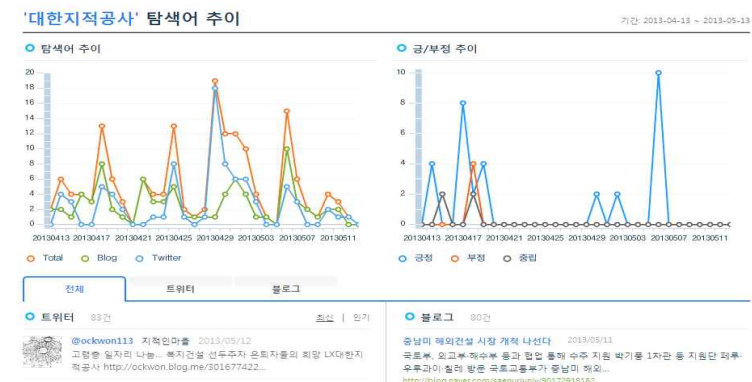
트위터 및 블로그의 검색 결과와 함께 관련어의 연관 키워드 순위, 감성 키워드 순위, 주간 급증 키워드 순위가 함께 표시된다. [그림 3.2]의 검색 키워드 순위로 살펴보면 『무료』, 담양지사가 어려운 군민을 위한 무료측량 실시로 지역민과 함께 하는 공사의 모범이 되고 있다는 내용과 『5월가정 달봉사활동』, 5월 가정의 달을 맞아 5월 11일(토), 안성시지사 직원들이 휴일도 반납한 채 보개면에 위치한 장애인 생활시설인 금관복지원을 방문해 가족처럼 행복한 사랑의 봉사활동을 펼쳤다는 내용이 순위에 있는 것으로 보아 긍정적인 키워드가 많은 것을 파악할 수 있다.

[그림 3.3] 소셜메트릭스 소셜 인사이트 탐색어 맵



소셜 인사이트 탐색어 맵은 탐색어의 연관어 및 카테고리별 분석을 해주는 기능을 가지고 있으며 『보건복지부』, 3년전 보건복지부가 대한지적공사의 항공사진과 지리정보시스템(GIS)을 이용해 조사한 결과 우리나라 묘지 면적이 여의도의 85배에 이르고 분포 수는 1435만 기(基)에 이르는 것으로 나타났다는 내용과 2013년 시니어 일자리 나눔 대회에서 은퇴자의 재취업을 적극 지원한 대한지적공사가 보건복지부 장관상을 수상했다는 내용이 주를 이룸에 따라 『보건복지부』와 『묘지』가 연관어 순위에 포함된 것을 알 수 있다.

[그림 3.4] 소셜메트릭스 소셜 인사이트 탐색어 추이



소셜 인사이트 탐색어 추이는 일정 기간 안에 소셜미디어의 종류별 탐색 결과와 관련 내용들을 분석하여 도출 된 긍정/부정/중립 결과를 그래프로 나타낸다.

[그림 3.5] 소셜메트릭스 소셜 인사이트 탐색어 여론



소셜 인사이트 부분에서 가장 중요한 탐색어 여론은 연관어의 긍정/부정/중립에 따른 탐색건수를 도표로 나타내어, 기업의 긍정적인 부분은 더욱 강조하여 부각되게 하고, 부정적인 여론은 검토하여 시정할 수 있도록 하는 일종의 보고서 역할을 한다.

소셜 메트릭스의 소셜 검색과 소셜 인사이트로 대한지적공사의 이미지를 분석하였다. 대체적으로 『수상실적』 및 『무료측량』, 『봉사활동』 등 긍정적인 내용이 많았으나 『불편』, 『애로』 등 부정적인 연관어도 발견되었다. 부정적인 연관어는 면밀한 검토를 거쳐 긍정적인 이미지로 만들어가야 할 기업의 숙제이다. 이제는 빅데이터를 적용하여 소셜미디어 뿐 만 아니라 바로처리콜센터, 접수자 및 측량팀의 상담정보, 전자민원, 청찬민원, CRM 상에서 나타나는 모든 고객의 목소리를 분석하여 고객의 니즈와 개선사항을 파악하여야 하며 이는 기업의 긍정적 이미지 제고에 큰 역할을 할 것이다.

4. 대한지적공사의 업무처리 현안

4. 1. 목표량 산정

목표량은 근시적으로는 한 해의 수익 성과를 가늠하며, 거시적으로는 공사의 경영 목표 설정에 영향을 주는 아주 중요한 요소이다. 이러한 목표량의 설정은 타당성은 물론 정확성을 갖춰 분석, 설계되어야 한다.

현재 우리 공사의 목표량 산정은 지사와 본부의 일반, 특수(개방업무 및 새업무), 부대 업무 목표량을 취합하여 이루어지고 있으며 이 중 전국에 배치 된 190여개 지사의 목표량이 가장 큰 비중을 차지한다. 각 지사의 목표량은 일반적으로 지사장 및 서무, 팀장들에 의해 직관적으로 산정되고 있다. 비중이 큰 일반, 특수 업무를 기준으로 보면 다년간의 완료량을 토대로 일정한 흐름에 맞춰 일반 업무의 목표량을 산정하며 팀마다 배정 된 개방업무 사업장의 수주 예상량을 팀장이 파악 또는 예측하여 특수 업무 목표량을 산정한다. 이렇게 산정된 목표량 중 분석에 의한 산정비율은 직관과 예측에 의한 산정비율에 비해 상당히 낮은 것이 사실이며 목표량 달성을 위한 인원 배정 및 사업효율 확대를 위해 보다 분석적인 목표량의 설정이 요구 된다.

4. 2. 민원처리

우리 공사는 한 근무지에서 일정기간 이상 근무를 하면 타 근무지로 이동을 하는 시스템을 가지고 있다. 여러 가지 취지가 있을 수 있겠으나 각 지사 및 지역에서 특성화 된 여러 가지 업무를 수행하며 많은 기술을 습득할 수 있다는 장점이 있는 것만은 분명하다. 그러나 이는 민원 발생 사유에서 만큼은 치명적인 약점으로 작용된다. 직원이 습득하고 있는 측량 성과 결정 방법이나 지역별 특성에 맞는 대처 방안 등 민원 예방, 해결을 위한 노하우가 인사발령 한 번에 날아가 버리는 것이다. 철저한 자료 조사를 바탕으로 민원을 예방하자는 정석에 가까운 해결 방안을 내놓을 수 있으나, 도면의 전산화 및 기술의 고도화를 바탕으로 진행되어온 도해지역의 측량성과 오류의 누적 문제는, 이론보다는 해당 근무지의 업무처리 노하우를 가진 개개인에 의해서 해결 되는 경우가 더 많은 것이 사실이다.

전산정보의 활용 또한 미미한 실정이다. 민원사항의 CRM, 자료부 등록을 시행하고 있으나, 이 시스템으로 인해 민원예방 및 해결에 직접적인 도움을 받는 경우는 극히

드물기 때문이다. 이에 대한지적공사는 고도화 된 정보화 시대를 맞아 측량 성과의 동일화하는 물론 고객의 목소리를 분석한 인사이트를 도출하여 민원 예방 및 해결에 적극 활용하여야 한다.

4. 3. 새업무 수주

우리 공사의 새업무 수주는 대부분 우리의 기술력으로 실현가능한 측량 및 행정업무를 관련부서 담당자와 협의하여 자발적으로 진행하여 왔다. 새업무 담당자의 아이디어와 직원들의 제안으로 매년 적지 않은 새업무 수익을 창출하고 있지만, 앞으로 관계기관 뿐 만 아니라 국민이 요구하는 서비스를 직접적으로 파악하여 우리의 업무에 적용시킨다면, 이는 제한된 환경에서 보다 넓은 사업영역 확보와 수입 증대를 가져올 것이다.

5. 빅데이터의 활용에 따른 기대효과 및 과제

5. 1. 빅데이터 활용 방안 및 기대효과

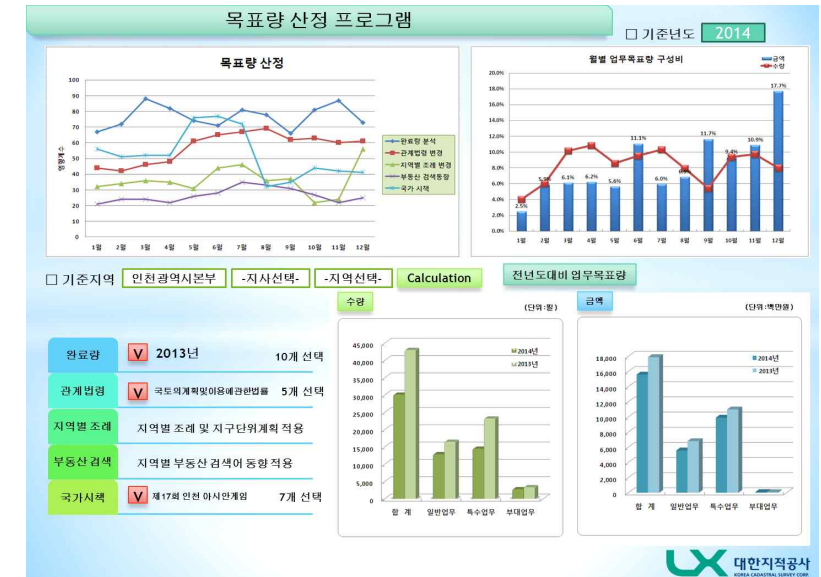
가. 목표량의 예측에 근거한 사업효율 증대 방안

빅데이터의 활용으로 분석에 근거한 정확한 목표량 산정 및 이에 따른 사업효율 증대를 가져올 수 있다. 장기간의 완료량 분석, 각종 관계법령의 변화와 지역별 조례 변경 등에 맞춰 분석한 1차적 목표량과 각종 포털사이트의 부동산 검색 동향, 국가시책 등에 따라 분석한 2차적 목표량을 합산하여 보다 정확한 목표량의 파악이 가능하다. 이 방안에 필요한 기술은 데이터 마이닝으로 변동요소들의 값을 정확히 파악하여 얼마나 정확한 함수를 만들어 낼 수 있는가 하는 것이 중요한 요소이며 단순하더라도 명확하게 목표량을 시각화 하는 기술 또한 필요하다. 분석에 필요한 요소들을 살펴보면 장기간의 완료량 분석은 현재에도 활용되고 있는 목표량 예측 방법으로서 다년간의 년·월별 업무 완료량을 토대로 계산되어지는 분석방법이다. 관계법령의 변화와 지역별 조례 변경에 따른 분석은 정확한 목표량 예측에 큰 비중을 차지한다. 지적측량업자의 업무 범위가 확장됨에 따라 발생하는 공사의 직접적인 수입 감소 요인 분석부터 건폐율·용적률의 변경, 개발제한구역의 지정, 지구단위계획 지정에 따른 최소대지 분할규모(규제) 등이 건축·개발 계획에 영향을 미쳐 이는 공사의 일반업무 중

가장 큰 비중을 차지하는 경계, 분할측량에 수입변화를 가져오기 때문이다. 포털사이트의 부동산 검색 동향 역시 목표량 변화를 예측할 수 있는 수단이다. 요즘은 부동산 중개업소에 들르지 않아도 실거래가 및 매물 정보를 수요, 공급자가 미리 알 수 있는 시대다. 토지의 매입을 위해서 수요자는 부동산관련 사이트에 해당 지역의 정보를 검색할 것이며 이를 카운팅하면 어느지역에 매매 붐이 이는지도 파악할 수 있다. 매매가 늘어나는 것은 곧 우리의 수입원인 경계, 분할측량이 늘어날 수 있는 조건이 되며 구글에서 독감관련 검색어 분석만으로도 독감 확산 초기 경보체계를 마련했듯이 검색어의 활용은 무엇보다 정확하고 신속한 예측을 가능하게 한다. 국가의 시책은 '2014년 제17회 인천 아시안게임 경기장 준공사업', '전철 개통사업', '경인아라뱃길 개통사업'등 특수한 사업의 진행이 공사의 수익에 영향을 미치는 시책으로 비정기적인 목표량인 하지만 해당 년도의 특수업무에 상당한 비중을 차지하는 업무를 말한다.

이러한 1, 2차적 목표량으로 산출된 정확한 목표량의 산정은 공사 내 인력 배치의 합리화를 가져와 사업효율을 증대 시킬 수 있으며 이는 지적제조사사업과 맞물려 탄력적 인력운영 및 사업전개에도 큰 힘으로 작용할 것이다.

[그림 5.1] 목표량 산정 프로그램 가상화면



나. 민원의 최소화를 통한 이미지 제고

민원의 발생은 대한지적공사에 막대한 손실을 가져온다. 민원의 종류에 따라 정신적, 물질적 손해는 물론 기업 이미지 실추라는 되돌리기 어려운 문제를 일으키기 때문이다. 더구나 직접적으로 재산권과 관련된 측량업무를 처리하고 있는 우리공사의 특성상 한 번 발생한 민원은 쉽게 해결하기가 어려우며 안타깝게도 2011년 11월 공영방송의 한 시사프로그램인 '추적60분'을 통해 독점에 대한 문제점이 방영되기까지 했다. 그렇다면 민원을 줄이고 기업의 브랜드 이미지를 높이기 위한 방법에는 어떤 것이 있을까. 바로 소셜미디어와 CRM 등의 활용이다.

좋은 예로 도로공사의 VOC 분석 시스템 구축사례를 들 수 있다. 도로공사 VOC 분석 시스템은 비정형 VOC 데이터를 통계, 분석 기반의 텍스트 마이닝 기술을 활용하여 상담 정보로부터 연관 정보 분석 및 지속적인 모니터링을 가능하게 하는 분석시스템으로 고객 니즈 발굴 및 개선 활동에 효과적으로 활용할 수 있는 시스템으로 개발 구축 되었다. 도로공사 VOC 분석 시스템은 콜 상담시스템의 콜 상담 정보, 민원 관리 시스템의 전자민원, 서면민원, 청찬민원, 고객 아이디어 그리고 채팅상담 시스템에 있는 채팅상담 정보들을 고도화된 언어처리 기법으로 분석하여 고객만족활동에 도움이 될 수 있는 각종 지표와 이슈들을 도출하고 활용할 수 있는 기능 제공을 목표로 성공적으로 구축되어 고객의 니즈 발굴 및 사업 개선에 적극 활용되고 있다.

[그림 5.2] 도로공사 고객의 소리 분석시스템



대한지적공사 역시 소셜미디어, 바로처리콜센터, 접수자 및 측량팀 상담정보, 전자민원, 청찬민원, CRM 상에서 나타나는 고객의 목소리를 분석하여 고객의 니즈와 개선사항을 파악하여야 하며 이를 성공적으로 구축, 활용 하였을 때 민원의 사전 예방 및 문제없는 사후 처리가 가능해져 기업의 이미지 제고는 물론 토지소유자 2/3이상의 동의를 받아 사업지구를 선정하는 지적재조사사업의 목표 달성률에도 큰 영향을 미칠 것이다.

다. 빅데이터를 활용한 수익 증대 방안

2004년 제11차 지적법 개정에 따라 지적측량업자는 경계점좌표등록부가 비치된 지역의 지적측량과 도시개발사업 등이 완료됨에 따라 실시하는 지적확정 측량을 수행할 수 있게 되었다. 었질러진 물을 다시 주워 담을 수 없듯 10년이 지난 지금, 지적측량업체는 예전과 비교할 수 없이 몸집을 키워왔으며 이제는 헌법소원을 통해 전면 개방을 요구 하며 공사에 큰 손실을 안기고 있다. 또한 각종 공공기관에서 발주된 지적측량업무의 대한지적공사 분담비율은 계속해서 하향 조정되고 있는 추세이며 추첨을 통해 탈락하기까지 한다. 공사 전 임직원의 노력에도 사회적 분위기는 우리에게 녹록치 않은 것만은 사실이다. 이러한 제한된 환경에서 수익을 증대시키는 방안이 바로 정보의 활용이다. 이는 두 가지로 나뉘볼 수 있다.

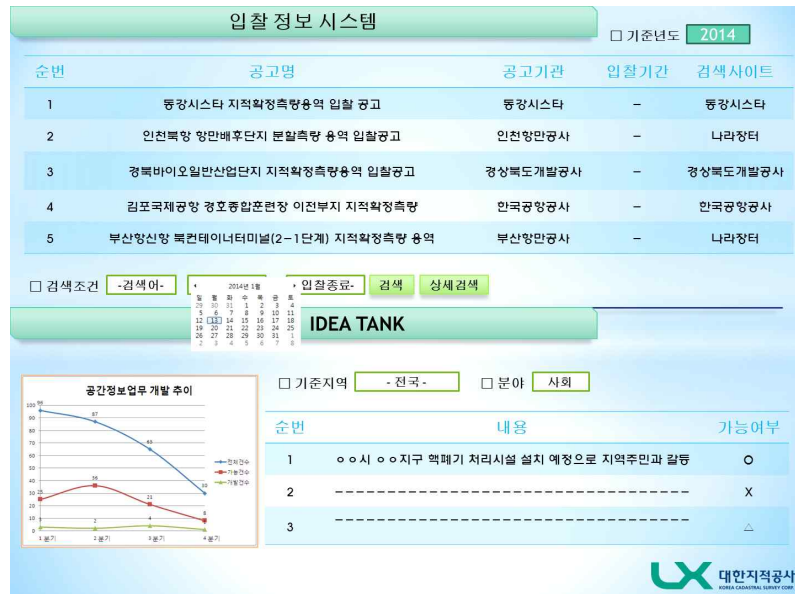
첫째, 실시간으로 수많은 공공기관 및 사업체 홈페이지의 고시, 공시란의 정보를 취득하여 활용하는 것이다. 직접적으로 측량신청이나 견적문의의뢰를 의뢰하지 않는 한 담당자가 모든 기관에서 접수되는 입찰정보를 확인하기에는 상당한 어려움이 있다. 웹 마이닝(Web Mining) 분석 기법을 활용해 업무수주가 가능한 공공기관 및 사업체 홈페이지의 고시, 공시, 게시판 정보를 크롤링(Crawling)²⁾하면 이는 서버에 실시간으로 축적되며 소규모의 분석인력을 가지고도 전국의 입찰 업무를 모니터링 할 수 있게 된다. 웹상의 데이터를 수집하는 데에는 기술적으로 데이터 수집자체를 막아놓은 경우를 제외하고 제한이 없으며 공정사용(Fair Use) 범조항으로 인해 불법이 아니다. 간과에 의한 업무 누수를 미연에 방지할 수 있으며 추후 해외사업 수주에 적용시킬 경우 막대한 수익 창출을 가져올 수 있을 것으로 기대 된다.

둘째, 빅데이터 분석을 통한 새업무 개발이다. 대한지적공사의 새업무는 대부분 우

2) 크롤링(Crawling) : 무수히 많은 컴퓨터에 분산 저장되어 있는 문서를 수집하여 검색 대상의 색인으로 포함시키는 기술로 요약하면 검색엔진의 로봇을 이용한 데이터 수집을 말한다. 어느 부류의 기술을 얼마나 빨리 검색 대상에 포함시키나 하는 것이 우위를 결정하는 요소로서 최근 웹 검색의 중요성에 따라 발전되고 있다.

리의 기술력으로 실현가능한 측량 및 행정업무를 관련부서 담당자와 협의하여 자의적으로 진행하여 왔다. 그러나 관계기관 뿐 만 아니라 국민이 요구하는 서비스를 직접적으로 파악하여 우리의 업무에 적용시킨다면 이는 제한된 환경에서 보다 넓은 사업영역 확보와 수입 증대를 가져올 것이다. 예를 들어 특정지역에 혐오, 기피, 오염시설이 들어서 있는 주민들이 이에 대한 불만을 가지고 있다는 데이터를 접하게 된다면 공사는 이러한 시설들과 피해 주거지역의 거리증명을 새업무로 발굴하여 공신력 있는 성과로 문제해결 방안을 마련할 수 있다. 또한 공사의 공간정보기술이 고도화됨에 따라 우리가 적용시킬 수 있는 새업무의 영역 역시 셀 수 없이 증가할 것으로 예상된다.

[그림 5.3] 입찰정보시스템 및 LX IDEA TANK 가상화면



5. 2. 빅데이터 활용 주요 사례

가. 샌프란시스코, 범죄 예방 시스템으로 안전한 지역사회 구축 사례

미국 샌프란시스코는 경찰청의 범죄 지도(crime map)를 기반으로 과거 8년간 범죄가 발생했던 지역과 유형을 세밀하게 분석하고 있는데, 이를 통해 샌프란시스코는 안전한 지역사회로 변화하고 있다. 여기에는 후속 범죄 가능성을 예측해 범죄를 사전에 예방하는 방식이 이용된다. 6개월간의 테스트 결과 예보 정확도가 71%에 달했으며 범죄가 예보된 10곳 중 7곳에서 실제 사건이 발생했다. 샌프란시스코는 이러한 범죄 지도를 구축함으로써 제한된 경찰 인력으로 광범위한 영역을 순찰하고 범죄를 예방하는 시스템으로 발전함으로써 안전한 지역사회를 구현할 수 있었다. 또한 과거 범죄자와 범죄 유형을 소셜 네트워크를 통해 지속적으로 관찰함으로써 그와 관련된 조직 및 범죄를 예방할 수 있었다.

이는 대한지적공사의 정확한 목표량 예측에 적용할 수 있는 유사 사례로 볼 수 있다. 다년간의 완료량과 관계법규, 지역별 조례 변화 등을 세밀하게 분석하여 추후 발생할 우리의 수입량을 예측할 수 있으며 여기에 2차적으로 변동량을 가지고 있는 검색어 동향 및 국가 정책을 분석하여 보다 정확한 목표량을 도출할 수 있다. 그러면 샌프란시스코 범죄 예방 시스템 구축 사례에서 볼 수 있듯이 제한된 인력으로 계획 업무 완료는 물론 사업 다각화 등 탄력적인 사업운영을 시도할 수 있을 것이다.

[표 5.4] 샌프란시스코 경찰청의 범죄 지도(Crime Map)



나. 유유계약의 ‘베노플러스’ 성공 사례

소셜 빅데이터 분석으로 10년간 평행선을 유지하던 연고제 판매량을 8개월 만에 46%나 늘린 제약회사가 있다. 유유계약이 그 주인공이며 회사는 특색 없던 마케팅을 소비자 행위와 심리 기반으로 변화시켜 이 같은 엄청난 성과를 거뒀다.

소셜 빅데이터 분석 대상이 된 제품은 2002년 출시된 베노플러스겔 연고다. 부은데·멍든데·타박상·벌레물린데 바르는 연고로 지금까지는 어린이를 대상으로 마케팅을 해왔다. 하지만 지난 10년간 매출은 제자리를 맴돌았으며 약가 인하로 전반적인 제약 경기가 나빠지면서 새로운 변화가 필요했다. 유유계약은 빅데이터 분석을 통한 새로운 마케팅을 도입하기로 하고 2012년 4월부터 다음소프트의 컨설팅을 받았다. 다음소프트는 트위터, 페이스북, 블로그 등 소셜 네트워크 사용자 인식을 분석했다. 그 결과 ‘붓기’나 ‘타박상’, ‘벌레물린데’와 비교할 때 ‘멍’에 대해서는 확실히 각인된 약이나 연고가 없었다. 오히려 계란이나 소고기, 찜질 등 민간요법이 더 많이 언급되고 있었다. 맨소래담이나 버즈비 같은 경쟁 제품보다 소비자 인식에 각인된 계란과 소고기를 이기는 게 시급했다. 소셜 네트워크 서비스 업체와 손잡고 ‘계란을 돌리는 것은 팔이 아프다.’ ‘소고기는 비싸고 비위생적이다.’라는 마케팅 메시지를 전하기 시작했다. ‘계란은 드세요. 멍은 베노플러스가 뺄게요.’라는 광고 메시지를 전파했다.

멍에 대한 소비자 트렌드 분석 결과 어린이보다는 성인 여성들이 더 많은 관심을 갖고 있었다. 이에 따라 여성 잡지에 광고를 내고 성형외과 홍보도 시작했다. 제품 포장도 ‘멍’이라는 단어가 맨 앞에 오도록 다시 디자인했다.

결과는 놀라웠다. 포털사이트에서 ‘멍 빨리 없애는 법’이라는 키워드는 전년 동기 대비 33% 감소한 반면, ‘베노플러스겔’은 557% 증가했다. ‘멍 빨리 없애는 법’이라는 검색 키워드 감소분이 베노플러스겔 검색으로 전이된 것이다. 2011년 4~12월 대비 2012년 4~12월 베노플러스겔 매출은 46%, 한해 매출은 62% 증가했다.³⁾

베노플러스 마케팅 성공 사례에서 우리가 주목하여야 할 부분은 분석을 통해 정확한 상황을 인지하였다는 것이다. 유유계약은 빅데이터 분석이 있기 전까지 멍에 대해 확실히 각인된 약이 없다는 정보를 알 수 없었으며 상식으로만 접근해 주 타겟층을 주의가 부족하여 다치기 쉬운 어린이로만 접근했다. 분석결과 베노플러스가 계란이나 소고기보다도 인지도가 낮아 이에 맞춤 마케팅이 필요하다는 정보를 얻을 수 있었으며 멍을 빼는데 관심이 많은 부류 역시 방학기간을 이용해 성형을 하거나 피부가 약한 성인 여성들이라는 것 또한 알 수 있었다. 이처럼 데이터의 분석은 기업의 올바른

3) 유유계약은 어떻게 베노플러스겔 매출을 46% 높였나. _ET News (2013년 1월 23일자)

목표설정 뿐 만 아니라 인지하지 못했던 부분에서의 성과창출까지 가능하게 한다. 지적측량이라는 테두리 안에서 공간정보업무로 새걸음을 내디딘 우리 공사가 새로운 공간정보업무영역을 구축하고 새업무를 개발해 나가기 위해 사회현상 및 이슈를 반드시 파악하고 분석해야만 하는 이유다.

다. 공공기관 빅데이터 활용 사례

우리나라에서는 국가정보화전략위원회가 빅데이터를 활용한 스마트정부 구현방안을 제시하는 등 국가 차원의 사업을 추진 중이며, ‘빅데이터 활용추진단’이 신설되어 빅데이터 경쟁력을 높이기 위한 국가·사회적 기반의 확보에 적극적으로 나설 계획이다.

【표 5.1】 공공기관 빅데이터 활용 사례

기 관	활용 사례
근로복지공단	△ 찾아가는 서비스를 통한 맞춤형 고객관계 관리 ◦ 전문적·단계적인 산재보상 서비스를 제공하기 위한 ‘찾아가는 서비스’도입 후 고객 만족도 향상 ◦ 전문적으로 구성된 찾아가는 서비스팀을 통해 환자 중심의 서비스 제공 가능 ◦ 찾아가는 서비스 도입 후 산재환자에 대한 초기 접근이 빨라져 업무상 재해를 결정하는 데 소요되던 기간이 19.3일에서 10.3일로 감소
통계청	△ 임금근로 일자리 통계로 일자리 현황 파악 지원 ◦ 청장년층의 취업난과 실업, 비정규직 고용불안 등 일자리에 대한 관심이 증가로 이에 맞춘 일자리 지표 마련 ◦ 고용보험, 국민연금, 건강보험 자료와 사업장 정보 파악을 위한 산재보험 자료를 활용하여 일자리 통계 지표 마련 ◦ 신규 일자리와 줄어드는 일자리가 즉각 파악되어 일자리의 안정성과 변동성을 쉽게 측정 할 수 있는 지표 마련
보건복지부	△ 지자체 공무원들의 복지 행정 처리를 지원하는 정보 시스템 ◦ 각종 사회복지 급여 및 서비스 지원 대상자의 자격 및 이력에 관한 정보를 통합관리 ◦ 복지대상자의 신청·접수·조사, 지원 여부결정, 급여지급, 사후관리 등 지자체의 복지 업무 처리 지원 ◦ 복지급여 지급과정에서 지급내역의 임의수정을 통한 부정 소지 차단 및 실명 확인 후 임금으로 제정의 투명성 제고 ◦ 1회 방문 신청 시 개인 및 가족의 복지 요구를 종합적으로 파악하여 가능한 모든 서비스를 안내, 상담할 수 있는 시스템으로 발전

5. 3. 빅데이터 활용 과제

가. 전문 인력 양성

빅데이터 기술 인프라 개발과 데이터 사이언티스트 인력양성은 빅데이터 시대를 앞서나갈 수 있는 필수적인 요건이다. 그러나 현재 데이터 사이언티스트의 역량을 갖춘 인력은 매우 부족한 실정으로 미국에서는 2018년까지 14만~19만명의 전문가와 150만명 정도의 데이터 관리자와 분석 인력이 부족할 것이라고 전망하였으며 EMC사의 글로벌 조사 결과 역시 향후 5년 동안 데이터 사이언티스트의 수요가 공급을 뛰어넘어 인재 부족이 심화될 것으로 예측되었다. 데이터 처리와 분석능력을 갖춘 인력은 IT 분야뿐 만 아니라 대부분의 기업과 조직에서 필수적으로 확보해야 할 핵심 인력으로 ①수학 능력 ②공학 능력 ③이해논리의 비판적 시각능력 ④글쓰기 능력 ⑤대화 능력 ⑥호기심과 개인의 행복을 기본 필요 조건으로 하고 있다.

또한 이런 능력을 갖춘 데이터 사이언티스트는 경영 전략과 밀접한 관계가 있으므로 외부에 아웃소싱 인력을 활용하는 것보다는 내재화가 바람직하다. 기업 내에서 사회 기회를 찾아내고, 전략적 통찰력을 발휘하기 위해서는 조직과 비즈니스 환경에 대한 깊은 이해가 필수적이기 때문이다. 대한지적공사 역시 조직의 민감한 데이터나 비공개 데이터 분석의 경우 외부 데이터 분석 인력을 활용하는 것은 위험성이 상존한다. 그리고 조직의 경쟁력 강화를 위해 조직 내부 인력들에게 데이터의 가치와 분석 역량을 키우는 사내 교육프로그램의 개발 또한 필요하다.

나. 개인정보 익명성 보장체계 확립

개인정보 익명성의 보장체계를 확립하기 위해서는 안전하고 신뢰할 수 있는 공공데이터 공개 및 활용을 위해 개인정보와 프라이버시 보호 관련 기본 원칙을 체계화하여야 한다. 목적명확화의 원칙, 수집제한의 원칙 등 개인정보보호법상의 원칙은 물론 빅데이터의 수집·이용·제공 등에 있어서 개인 식별 정보의 변환 및 삭제의 원칙, 데이터 원본보다는 집계 값의 활용 원칙 등 공공데이터 활용에 따른 국민 불안 해소를 위해 공공데이터 활용 시 개인정보침해 최소화 장치를 마련해야한다. 또한 민감한 개인정보에 대해 정보보안 수준별 권한 및 관리체계를 명확히 제시하여야 한다.

그리고 빅데이터의 안전한 공유와 유통을 위한 강화된 보안대책을 마련한다. 불필요한 데이터 사용을 원천적으로 차단하고, 빅데이터 불법 수집 및 오남용 방지를 위

한 규정을 도입 하여야 하며 대한지적공사내 전문기술과 경험을 갖춘 보안전담조직의 설치·운영으로 보안성 모니터링체계를 상시화 하는 것 역시 필요하다.

마지막으로 데이터의 개방·공유·활용에 따른 공공데이터 활용 가이드라인을 마련해야한다. 개인정보가 노출되지 않도록 개인식별 속성 제거, 암호화 등에 관한 원칙을 수립하는 것이 필요하며, 소셜 데이터를 연계하여 개방·공유·활용 등의 경우는 원본 데이터, 위치정보 등에 대한 비저장 및 개인 비식별 등의 원칙을 마련하여야 한다.

3. 결 론

현재의 대한지적공사는 빠르게 변화하는 정보화 시대에 발맞춰 적극적으로 대응하고 있다. COS, MOS, ERP, CRM 등의 시스템으로 보다 유연하고 효율적인 업무 처리가 가능해 졌고 이는 공사의 업무처리환경에 시간 절약과 편리함이라는 큰 장점을 가져온 것이다.

그러나 이제 데이터의 축적만으로는 경쟁할 수 없는 빅데이터의 시대다. 기업의 IT기술을 전자화·자동화의 단계에서 멈춘 사업자와, 사업에 도움이 되는 인사이트 도출의 단계로 발을 내디딘 사업자는 미래사회에서 큰 경쟁력의 차이를 보일 것이다. 대한지적공사 역시 제한된 지적축량 시장안에서 고객의 행동을 미리 예측하고 대처 방안을 마련하여 기업 경쟁력을 강화시키고, 생산성 향상과 비즈니스 혁신을 가능하게 하기 위해선 잘 정제된 대한지적공사 내부의 축적자원 뿐 아니라 무수히 많은 외부의 정형/비정형 데이터를 분석하여 가치 있는 데이터로 발전, 활용시킬 필요가 있다. 이러한 빅데이터를 활용하기 위해 공사에 필요한 분석 기법은 기존의 데이터를 활용하고 웹상의 정보를 분석, 처리하며 여론의 목소리를 도식화 할 수 있는 데이터 마이닝, 텍스트 마이닝, 웹 마이닝, 소셜 마이닝 기법이며 분석 기술은 앞서 소개한 하둡(Hadoop), R, NoSQL 모두 이다. 이 세가지 기술은 글로벌 서비스에 적합한 규모 확장성과 비용 효율성을 갖춘 오픈소스 기술들로 공사내 구축된 RDBMS 시스템과의 유연성과 확장성을 고려하여 설계 되는 것이 바람직하다.

빅데이터의 구축에 따른 활용으로 얻을 수 있는 기대효과는 첫째, 목표량의 예측에 근거한 사업효율 증대 방안, 둘째, 민원의 최소화를 통한 이미지 제고, 셋째, 빅데이터를 활용한 수익 증대 방안을 들 수 있다. 급속성장의 중심에 서있는 IT기술의 특성상 미래에 나타날 빅데이터 신기술의 범위와 생성 주기를 예측하는 것은 상당한 어려움이 있다. 그러나 이는 반대로 대한지적공사에서 적용 가능한 획기적인 성과창출방안 역시 무궁무진하다는 것을 뜻한다.

이러한 빅데이터를 효율적으로 활용하기 위해 공사는 경영을 위한 전략적 통찰력과 데이터의 가치 분석 역량을 가진 데이터 사이언티스트의 육성이 필요하며 빅데이터의 불법 수집 및 오남용 방지를 위한 규정 도입과 공공데이터 활용 시 개인정보침해의 최소화 장치를 마련해야한다.

빅데이터는 생산성 향상, 정부 혁신에 기여할 뿐 만 아니라 지금 준비하지 않으면 그 격차를 따라잡기 어려운 미래 경쟁력의 핵심 수단이다. 타기업, 기관의 성공사례를 분석하여 우리 공사에 적용 가능한 최적의 빅데이터 분석 인프라를 구축시킬 필요가 있으며 이는 현재는 물론 다가올 미래에 공사의 가장 큰 자산이 될 것이다.

참 고 문 헌

I. 단행본

- i. 함유근, 채승병, 『빅 데이터, 경영을바꾸다』 (삼성경제연구소, 2012)
- ii. 송길영, 『여기에 당신의 욕망이 보인다(빅데이터가 찾아낸 70억 욕망의 지도)』 (쌤앤파커스, 2012)
- iii. 송민정, 『빅 데이터가 만드는 비즈니스 미래지도』 (한스미디어, 2012)
- iv. 스키 리스캐, 『빅 데이터 비즈니스』 (도서출판 더 숲, 2012)

II. 논문

- i. 김지숙, “빅데이터 활용과 분석기술 고찰”, 고려대학교 대학원 석사학위 논문, 2012.
- ii. 박준규, “빅데이터를 위한 분석기술 활용방안 연구”, 세종대학교 대학원 석사학위 논문, 2012.

III. 기타자료

- i. 김동한, “빅 데이터의 핵심 플랫폼, 기업용 하둡 동향”, 『정보통신산업진흥원 (주간기술동향)』, 2013. 2. 13.
- ii. 정지선, “성공적인 빅데이터 활용을 위한 3대 요소 : 자원, 기술, 인력”, 『한국정보화진흥원(IT & Future Strategy)』, 2012. 4. 12.
- iii. 정지선, “新가치창출 엔진, 빅 데이터의 새로운 가능성과 대응 전략”, 『한국정보화진흥원(IT & Future Strategy)』, 2011. 12. 30.
- iv. 조성우, “Big Data 시대의 기술”, 『KT경제경영연구소』, 2011. 10. 5.