

빅데이터 정보시스템의 구축 및 사례에 관한 연구

(A Study of Big Data Information Systems Building and Cases)

이충권*

(Choong Kwon Lee)

요약

빅데이터와 관련하여 많은 성공사례들이 보고되었지만 실제로 시스템을 구축하는 데 있어서는 여전히 많은 어려움이 있다. 기술적인 측면에서 데이터의 수집과 저장, 처리와 분석, 그리고 표현과 사용에 이르는 전 과정을 포괄적으로 이해해야 하고, 비즈니스적 측면에서는 구축된 시스템으로부터 얻을 수 있는 가치를 미리 파악하여 투자를 감행해야 하는 경영진에게 설명해야 한다. 본 연구는 빅데이터 정보 시스템을 구축하는 것과 관련된 사항들을 쉽게 파악할 수 있는 5W 1H 프레임워크를 제공하고, 제시된 프레임워크를 기존의 빅데이터 사례들에 적용한 예시를 보여주었다. 투자를 위한 경영진의 의사결정을 이끌어내고 빅데이터 프로젝트의 종합적인 이해와 관리에 도움을 줄 수 있을 것으로 기대된다.

■ 중심어 : 빅데이터 ; 정보시스템 ; 5W 1H ; 사례연구

Abstract

Although many successful cases regarding big data have been reported, building information systems of big data is still difficult. From the perspective of technology the builders need to understand the whole process of systems development ranging from collecting, storing, processing, and analyzing data to presenting and using information. Whereas, from the perspective of business, the builders need to understand the values of the proposed big data project and explain to top managers who have to make a decision of the risky investment. This study proposes a framework of 5W 1H that can help the builder understand things related to the development of big data information systems. In addition, big data cases from the real world have been illustrated by applying to the framework. It is expected to help builders understand and manage big data projects and lead managers to make better decisions of the investment to the development of information systems

■ keywords : Big data ; Information systems ; 5W 1H ; Case studies

I. 서론

데이터는 현실 세계로부터 관찰이나 측정을 통하여 수집한 사실을 문자나 기호, 그림 등을 이용하여 기록한 것이다. 개인은 생활을 기록하고, 국가는 행정을 기록하고, 기업은 비즈니스 활동을 기록한다. 기록된 데이터를 기반으로 의사결정에 영향을 끼치거나 도움이 되는 정보를 생성한다. 모든 정보는 데이터로부터 나오고, 정확한

데이터만 있으면 정보는 얼마든지 만들 수 있다.

조직의 내부 뿐만 아니라 외부에서 발생하는 많은 양의 데이터를 수집하여 저장하고 분석하여 의사결정에 도움이 되는 정보를 얻어내려는 시도가 여러 분야에서 이루어지고 있다. 일반인들이 주고받는 문자메시지나 전기 및 수도 사용량, 신문기사 등과 같은 무한대에 가까운 데이터를 분석하여 의사결정에 활용하여 성공한 사례들이 보고되고 있다. 막강해진 컴퓨터 처리능력을 활용하여 조직의 내외부에서 발생하는 많은 양의 데이터를 수집하고 분석하여 현상에 대한 더 나은 이해와 조직의 의사결

* 계명대학교 경영정보학과 부교수

정에 활용하는 것을 빅데이터라고 한다[1].

다양한 분야에서 보고되는 성공사례에도 불구하고, 빅데이터 정보시스템의 가치를 경영진에게 설명하고 구축을 위한 투자를 결정하도록 하는 것은 쉽지 않은 일이다[2]. 조직에서 정보시스템을 개발하고 운영하는 역할을 맡은 부서에서 빅데이터의 수집과 활용에 이르는 과정과 구축된 시스템으로부터 얻을 수 있는 가치를 기존의 사례들을 기반으로 이해할 필요가 있다.

본 연구는 빅데이터의 수집과 저장, 처리와 분석, 그리고 표현과 사용에 이르는 과정을 포괄하는 정보시스템을 구축함에 있어서 고려해야 할 사항들을 파악하는데 도움을 되는 프레임워크를 제시하고자 한다. 또한, 제시된 프레임워크의 관점에서 빅데이터와 관련된 다양한 사례들을 살펴봄으로써 빅데이터 정보시스템의 구축에 있어서 필요한 사항들에 대한 종합적인 이해를 돕고자 한다.

II. 본 론

1. 빅데이터의 개념

Laney[3]는 인터넷을 기반으로 한 전자거래의 발달이 가져올 새로운 기회와 도전의 중심에는 데이터가 있고, 그 특성이 3V라고 언급하면서 양(Volume)과 속도(Velocity), 그리고 다양성(Variety)이라고 하였다. 양은 기술과 비용의 문제로 과거에는 디지털화 하지 못 했던 데이터를 정보 기술을 통하여 수집하고 저장할 수 있게 됨으로써 크다는 뜻을 가진 빅(big)이라는 단어를 덧붙여야 할 정도로 폭발적으로 증가한 것을 의미한다. 속도는 컴퓨터의 처리장치가 발달하면서 데이터가 생성되고 분석되는 시간이 매우 빨라진다는 것이다. 인터넷은 공간적 제약을 극복하여 전 세계에서 발생하는 데이터를 거의 실시간으로 수집하고 처리할 수 있는 기술적 인프라가 되었다. 다양성은 미리 정해진 형태로 발생하는 데이터 뿐만 아니라 일반인들이 일상생활에서 생성하는 텍스트, 음성, 그리고 영상 등과 같은 비정형 데이터도 수집과 분석의 대상이 된다는 것을 말한다.

빅데이터는 보편적으로 사용되는 소프트웨어 도구들의 능력범위를 벗어난 크기의 데이터를 제한된 시간 내에 수집하고 처리해야 한다[4]. 데이터의 양이 급격히 커지면서 2000년대에는 테라바이트(terabyte) 정도를 빅데이터라고 하였지만, 2010년대에 들어서면서 페타바이트(petabyte)나 엑사바이트(exabyte)로 빅데이터의 개념이 확대되고 있다. 이처럼 양적인 개념에서 출발한 빅데이터가 최근에는 질적인 의미를 중요시하는 방향으로 나아가고 있다. Hashe m 등[5]은 복잡하고 다양한 형태로 저장된 대량의 데이터로부터 숨어 있는 가치를 찾아내는 기술이 진정한 빅데이터라고 하였다. 빅데이터는 모집단으로부터 표본을 추출하지 않고 현장에서 발생하는 데이터를 실시간으로 모아서

현상에 대한 이해와 판단에 도움이 되는 정보를 생성함으로써 의사결정적 가치를 제공하는 기술인 것이다. Laney [3]가 제시한 빅데이터의 특성인 양, 속도, 그리고 다양성이라는 3V에 가치(value)를 더하여 4V로 확장되어 발전하고 있는 것이다.

2. 빅데이터 정보시스템

빅데이터의 등장과 발전에 있어서 하드웨어와 소프트웨어의 성능과 가격이 급격히 낮아진 것은 중요한 요인이었다. 분산 및 병렬처리와 하둡(Hadoop), 그리고 클라우드 컴퓨팅과 같은 기술적 발전은 데이터의 수집과 저장에 있어서 획기적인 변화를 가져왔다. 그러나, 이러한 기술적 변화가 단순히 개인의 생활이나 기업의 활동을 편리하게 하는 수준에서 머물렀다면 정보기술 투자의 효과는 제한적이라고 할 수 밖에 없을 것이다. 실제로 정보기술 투자로 인한 효과에 관해서 의문이 제기되어 왔는데, '정보기술 생산성의 역설[6, 7]'로 표현되거나 Carr[8]는 정보기술은 아무런 도움도 되지 않는다는 주장을 하기도 하였다. 이러한 주장의 배경에는 정보기술에 대한 투자가 의사결정의 질을 높여서 생산성에 기여하기 보다는 단순한 업무지원의 역할 밖에는 하지 못한다는 것과 구축된 시스템이 기계적인 전산화에 그쳐서 새로운 가치의 창출로 이어지지 못한다는 비판의 의미를 담고 있었다.

조직의 활동에서 발생하는 각종 데이터를 수집하고 처리하여 의사결정을 위한 정보를 생성할 수 있도록 시스템을 구축하는 것은 정보기술에 대한 투자효과를 높이는 데 있어서 필수적이다. 이처럼 오랫동안 데이터는 조직 내부에서 발생하는 데이터를 수집대상으로 하였으나, 정보기술의 발달로 시민들의 공과금 데이터나 소셜네트워크에서 발생하는 문자나 사진 등을 수집하여 가치 있는 정보를 생성하는 사례가 나타나고 있다. 이것은 정보시스템의 필수적인 입력요소인 데이터를 조직 내부 뿐만 아니라 외부의 다양한 소스로 부터 확보하여 보다 빠르고 정확한 의사결정에 도움을 줌으로써 정보기술에 대한 투자가 더욱 효과적일 수 있게 되었음을 의미한다.

예를 들어, 어떤 상품에 대한 일반인들의 초기반응을 알아보기 위하여 트위터 문자를 분석하거나 고객과의 통화 내용을 녹화하여 데이터베이스에 저장하고 분석하여 문제점을 파악할 수도 있다. 이처럼 빅데이터 정보시스템은 기업의 내외부에서 대량으로 발생하는 데이터를 수집하고 처리하여 지속적으로 정보를 제공할 수 있는 시스템을 말한다. 빅데이터를 활용한 다양한 사례들이 보고되었지만, 기업이나 기관의 입장에서 안정된 데이터 원천과 정보의 생성을 위한 시스템을 구축하려면 어떻게 해야 하는지에 관한 논의는 부족하였다. 따라서, 본 연구는 기존의 빅데이터 사례들을 정보시스템 구축의 관점에서 살펴보고자 한

다. 이러한 연구는 어떤 조직의 정보기술에 대한 투자가 단순히 기술 그 자체의 도입으로 인한 단기적인 효과에 그치지 않고, 조직의 의사결정과 활동에 도움이 되는 정확하고 빠른 정보를 지속적으로 제공할 수 있는 시스템의 구축을 통하여 조직의 성과에 연결될 수 있는 방법을 찾는 데 도움을 줄 것이다.

3. 빅데이터 정보시스템 구축을 위한 5W 1H

기자들이 사실과 사건에 관한 기사를 작성하는데 있어서 누가(Who), 언제(When), 어디서(Where), 무엇을(What), 왜(Why), 그리고 어떻게(How)를 의미하는 5W 1H는 미디어 저널리즘 분야에서 반드시 지켜야 할 원칙처럼 인식되어왔다. 기사 내용은 전달하고자 하는 정보를 담고 있는데, 그 정보는 여러 가지 데이터들을 조합하여 만들어진다. 따라서, 기사 속의 정보를 생성하는데 필요한 데이터는 모두 5W 1H에 근거하여 수집된다. 빅데이터 정보시스템도 다양한 소스로부터 많은 데이터를 수집하여 정보를 생성한다는 관점에서 바라보면 저널리즘의 5W 1H 프레임워크는 일맥상통하는 면이 있다. 표 1은 빅데이터 정보시스템의 관점에서 5W 1H 프레임워크를 보여주고 있다.

표 1. 빅데이터 정보시스템의 5W 1H 프레임워크

	데이터 ⇔					정보
	수집	저장	처리	분석	표현	
Who	○					○
When	○		○	○	○	○
Where	○	○	○		○	○
What	○		○		○	○
Why	○					○
How	○	○	○	○	○	○

가. 수집

빅데이터 정보시스템을 구축함에 있어서 데이터의 수집은 전체 시스템의 가장 기초가 되는 영역이다. 데이터를 누가(Who), 언제(When), 어디서(Where), 어떤 유형(What type)으로 발생시키는지 조사하고 파악해야 한다. George 등[9]은 빅데이터의 소스를 5가지로 구분하였는데, 정부에 의한 공공(public) 데이터, 개인들에 의해 생성되는 프라이빗(private) 데이터, 인터넷사용으로 인해 발생하는 접속로그와 같은 배출(exhaust) 데이터, 소셜미디어에서 생성되는 커뮤니티 데이터, 그리고 개인의 행동에 의해 만들어

지는 자기계량(self-quantification) 데이터가 있다.

기술의 발전으로 데이터의 수집이 용이해졌지만, 수집해야 할 데이터의 양과 질, 방법과 목적, 그리고 비용과 시간 등을 고려해야 한다. 수집해야 할 데이터의 양(How much)은 사용목적(Why)과 관련이 있다. 사용목적과 관련이 없는 불필요한 데이터를 수집하는 것은 비용과 시간에 있어서 낭비를 초래하기 때문이다.

나. 저장

다양한 유형으로 생성되는 대량의 데이터를 어디에(Where) 어떻게(How) 저장할 것인가의 문제는 다음 단계인 처리와 분석과 관련이 있다. 체계적으로 저장하지 않으면 처리와 분석이 용이하지 않기 때문이다. 특히, 텍스트, 음성, 영상 등과 같은 비정형 데이터는 저장방식에 따라서 처리와 분석에 필요한 시간이 크게 달라지기 때문에 정보를 얻기에 용이한 형태로 저장을 해야 한다.

기술적 측면에서 오픈소스 하둡(Hadoop)은 빅데이터의 저장에 있어서 사실상 표준이 되었고, 많은 업체들이 하둡과 연동되는 솔루션을 출시하고 있지만, 빅데이터의 저장과 관련해서는 여전히 고비용을 요구한다. 특히, 빠르게 누적되는 빅데이터의 속성상 용량의 한계를 극복하기 위한 확장성도 중요한 고려대상이다.

다. 처리

처리는 빅데이터를 분류와 여과를 통하여 분석이 가능한 상태로 바꾸는 것을 말한다. 분석 전에 실행되어진다고 하여 전처리라고도 부르는데 스토리지 비용을 줄이려면 저장단계 전에도 부분적으로 수행해야 한다. 데이터의 내용을 어떻게(How) 정제하고 분류하여 분석과 관련 없는 부분들을 제거할 것인지 그 방법과 과정을 미리 설계해야 한다. 이러한 처리과정을 통하여 데이터의 축소와 통합, 그리고 변환이 일어나는데[10], 처리과정은 빅데이터 정보시스템에서 가장 많은 시간을 필요로 하는 단계로 알려져 있다[11]. 데이터가 클수록 시간을 단축시키려면 분산, 병렬 처리를 해야 한다. 데이터를 저장과 동시에 실시간(When)으로 처리할 것인지, 아니면 저장해 두었다가 나중에 처리할 것인지도 고려해야 한다.

라. 분석

빅데이터는 다수의 샘플을 기반으로 하기 때문에 정보의 왜곡이 적을 뿐만 아니라 변수 선택의 폭을 넓힐 수 있고 빠르게 최신의 결과를 확보할 수 있다는 것이 기술적 장점이다. 즉, 빅데이터 정보시스템은 기업내부에 저장된 거래기록

을 바탕으로 고객의 특성이거나 구매패턴을 이해할 수 있을 뿐만 아니라 기업 외부에 있는 날씨, 교통, SNS 등으로부터 얻은 데이터를 활용하여 인과관계의 파악이나 예측도 가능하다.

그러나, 빅데이터 정보시스템에서 데이터 자체 보다는 이를 활용하여 새로운 인사이트를 도출하는 분석을 통해 비즈니스적 이익을 창출하는 것이 중요하다[12]. 분석의 결과로 가치있는 정보를 얻을 수 있다는 확신이 있어야 투자가 가능하기 때문이다. 빅데이터를 활용하여 의사결정의 질을 높이고 새로운 가치를 창출하려면 실제 일어난 현상에 대한 이해 뿐만 아니라 미래에 대한 예측까지도 가능한 분석을 해야 한다. 저장되고 처리된 데이터를 언제(When), 어떻게(How) 분석하여 제공할 것인가는 빅데이터 정보시스템의 구축에 있어서 가져야 할 중요한 질문이다. 가치 있는 정보를 제공하려면 분석에 필요한 시간과 방법을 미리 정해 놓아야 하는 것이다.

마. 표현

분석을 통하여 얻어진 어떤(What) 정보를 언제(When), 어떻게(How) 보여주느냐에 따라서 사용자의 이해 정도가 달라질 수 있다. 도표나 그림을 통하여 요약된 정보를 제공하여 직관적으로 이해하도록 만들 수도 있고, 요약된 정보의 구체적인 세부내용을 확인할 수 있도록 드릴다운 메뉴를 제공할 수도 있다. 최근에는 데스크탑이나 노트북 컴퓨터 뿐만 아니라 태블릿이나 스마트폰 등과 같은 인터페이스도 고려 대상이 되었다. 스마트폰의 성능향상으로 앱을 개발하여 보여줄 수도 있고, 단순히 그래픽 문자메시지로 보내줄 수도 있다. 이처럼 사용자가 기기에 상관없이 어디서나(Where) 제공된 정보를 볼 수 있도록 하는 ‘One Source Multi Use’를 구현하기 위한 기술적 요구사항들을 소프트웨어 개발 시에 고려해야 한다.

바. 사용

빅데이터로부터 얻은 정보를 사용하는 것은 시스템의 성공여부를 좌우하는 결정적인 요인이다. 좋은 정보를 제공하고 성능이 우수한 시스템이라고 하더라도 사용되지 않는다면 결국은 실패로 귀결될 것이다. 따라서, 빅데이터 정보시스템을 구축하기 전에 누가(Who), 언제(When), 어디서(Where), 무슨(What) 정보를 어떻게(How), 왜(Why) 사용하는지를 조사하고 파악해야 한다. 시스템의 구축 후에는 주로 사용하는 정보와 사용자, 접속 방법과 시간, 장소, 목적 등과 같은 자료를 확보하여 분석하여 시스템의 운영과 개선에 반영해야 한다. 특히, 빅데이터의 사용목적(Why)을 명확히 함으로써 구축될 시스템의 가치를 설명할 수 있어야 한다. 예를 들어, 고객만족이 목적이라면 데이터의 수집과 분

석을 통해 고객에게 어떤 이익을 가져다 줄 수 있는지를 명확히 해야 정보시스템 구축을 위한 투자결정을 할 수 있는 것이다.

III. 사례분석

1. 공공부문

정부와 공공기관은 각종 행정서비스를 제공하기 위하여 시민들로 부터 많은 데이터를 수집하고 있다. 전구나 수도와 같은 도시 인프라로 부터 나오는 데이터의 패턴을 분석하여 누전이나 누수를 탐지하고 사용량을 예측하기도 한다. 이처럼 일반인들이 사용하는 공공서비스들로부터 데이터를 수집하여 서비스 품질의 향상이나 고객만족을 달성할 수 있다.

시스코[13]는 영국 가정의 전기공급에 있어서 70%에 가까운 비중을 차지하고 있는 센트리카를 빅데이터의 성공사례로 들고 있다. 스마트 계량기 350만여 대를 무료로 배포하고 사용량을 포함하여 기온, 습도 등과 같은 날씨 데이터를 30분 단위로 수집하고 분석함으로써 사용패턴을 파악하고 가정의 에너지 절약을 도와주고 전력수요를 예측하여 미래의 전력소비에 대비한다. 표 2는 센트리카의 빅데이터 정보시스템 사례를 5W 1H 프레임워크에 근거하여 요약된 정보를 보여주고 있다.

표 2. 센트리카 빅데이터 정보시스템 기본 고려사항 예시

	수집	저장	처리	분석	표현	사용
Who	시민					고객팀
When	매 30분		30분	매일	매일	일/월
Where	가정	스토리지 X	지역별		데스크탑	지역별 패턴
What	사용량	수치 텍스트	SAP HANA		비교	소비 예측
Why						에너지 절약
How	연간 1.2 terabyte		인메모리 DB	패턴	그래프	연간 190포인트 절약

2. 제조업 부문

제조업체들은 부품구매에서 완성품 생산에 이르는 생산공정에서 발생하는 데이터를 수집하고 관리하여 원가를 절감하거나 제품을 혁신해야 한다. 복잡한 공정에서 생성되는 정

형/비정형 데이터는 센서나 계측기를 통하여 얻어지는데, 이상징후를 실시간으로 감지하여 원인을 찾아내거나 불량품의 생산비용, 기계의 잔여수명 등을 예측할 수 있다[14].

Jacobs[15]는 지멘스가 암베르크 공장에서 2억 유로에 달하는 공장자동화시스템을 신규로 투자하여 구축하면서 빅데이터 개념을 활용하여 생산성은 8배 증가시킨 반면에 불량률은 100만개 중 11.5개 수준으로 줄었다. 매일 5000만 건의 데이터를 실시간으로 수집하여 문제가 발생한 공정은 담당자에게 알려주고 원격제어를 통해 문제를 해결하고 공정의 최적화를 추구한다. 일정 시간 이상의 휴식이 예상되는 기계들의 전원은 자동으로 차단되도록 하여 에너지 효율을 높일 수 있다. 제품의 기획부터 판매 이후까지 모든 데이터를 수집해서 얻은 정보는 제품의 개선이나 개발에 반영한다. 표 3은 지멘스의 빅데이터 정보시스템 사례를 5W 1H 프레임워크에 적용하여 보여준다.

표 3. 지멘스 빅데이터 정보시스템 기본 고려사항 예시

	수집	저장	처리	분석	표현	사용
Who	기계					공장관리팀 제품개발팀
When	실시간		수시	수시	수시	수시
Where	공장	스토리지 X	기계별		데스크탑	기계별 작동패턴
What	작동	수치 텍스트	PLM Software		비교	오류탐색
Why						공정관리
How	매일 5천만 건		원격 제어	패턴	이메일	불량률 감소

3. 서비스업 부문

서비스업은 고객의 신상정보나 거래기록을 수집, 분석하여 새로운 가치를 창출하려 한다. 미국의 세튼 헬스케어 패밀리 병원의 사례는 의료기관에서 환자와 진료기록을 기반으로 증세와 질환의 패턴을 발견함으로써 장기간 축적된 데이터로부터 과거와 현재의 규칙성과 상관관계를 밝혔다[16]. 스페인의 의류업체인 ZARA는 전 세계에 널려 있는 매장들의 판매실적을 실시간으로 분석하여 매장과 품목별로 수요와 판매량을 예측하여 적정재고를 제공하는 정보시스템을 구축하였다[17].

최근에는 이미 판매된 제품의 사후관리가 중요해지면서 구매고객으로부터 데이터를 얻어서 제품에 관한 정보를 얻으려는 시도도 나타났다. 예를 들어, 삼성전자나 LG전자, 그

리고 현대자동차 등과 같은 기업들은 트위터나 블로그, 웹페이지에 올라오는 비정형텍스트 데이터를 수집하고 분석하여 제품에 대한 소비자들의 만족이나 불만 사항들을 수집하고 분석하여 서비스나 제품의 개선에 활용한다[18, 19]. 또한, 일반인들의 소셜네트워크 데이터를 분석하여 영화의 성공여부를 미리 예측하거나 마케팅에 활용하고 있다[20, 21].

Henschen[22]은 전 세계 4만 개 이상의 매장에서 매주 2억 5천만 건 이상의 거래로부터 2.5 페타바이트의 데이터를 수집, 분석하여 고객별로 맞춤형 정보를 제공하는 카탈리나마케팅사의 사례를 보고하였다. 고객의 구매상품, 수량, 빈도, 액수 등과 같은 거래데이터를 분석하여 거래발생 시에 즉석에서 맞춤형으로 할인쿠폰을 제공한다. 어떤 고객의 거래가 쌓일수록 해당 고객에게 더욱 정확한 마케팅정보를 제공함으로써 재방문과 추가구매를 통하여 매출을 증대할 수 있다. 표 4는 카탈리나마케팅의 빅데이터 정보시스템 사례를 5W 1H 프레임워크에 적용하여 보여준다.

표 4. 카탈리나의 빅데이터 정보시스템 기본 고려사항 예시

	수집	저장	처리	분석	표현	사용
Who	고객					
When	거래발생 시		매주	매주	수시	수시
Where	매장	스토리지 X	매장별		매장 모니터	
What	거래	수치 텍스트		구매 기록	할인율	쿠폰발행
Why						고객만족
How	2.5 petabyte			패턴	종이 / 모바일	매출액 증대

IV. 결 론

빅데이터에 관한 많은 성공사례들이 보고되었지만 개별 기업이나 기관에서 대규모 데이터를 수집하고 분석하기 위한 정보시스템을 구축하는 데는 여전히 많은 장애물들이 있다. 데이터 수집과 저장을 위한 데이터베이스와 스토리지, 분석을 위한 알고리즘과 소프트웨어의 도입 및 개발, 정보의 표현에 있어서 직관적인 판단이 가능할 정도의 시각화, 그리고 사용자들의 접속과 정보활용을 지속적으로 모니터링하고 유지보수하는 작업은 대규모 투자와 관리가 필요하다. 이러한 투자와 관리를 원활하게 진행하려면 빅데이터 정보시스템을 구성하는 데이터 수집과 저장, 처리와 분석, 그리고 표

현과 사용에 이르는 과정을 전체적으로 이해하고 관련자들에게 설명할 수 있어야 한다.

본 연구는 빅데이터 프로젝트를 추진해야 할 담당자가 정보시스템의 구축과 관련된 사항들을 일목요연하게 정리하고 쉽게 이해할 수 있는 5W 1H 프레임워크를 제시하였다. 시스템 구축 담당자가 필요사항들을 파악할 수 있도록 제시된 프레임워크를 기존의 빅데이터 사례들에 적용하여 예시를 보여주었다. 투자를 위한 경영진의 의사결정을 이끌어내고 빅데이터 프로젝트의 종합적인 이해와 관리에 도움을 줄 수 있을 것으로 기대된다.

REFERENCES

- [1] A. Jacobs, "The pathologies of big data," *Communications of the ACM*, vol. 52, no. 8, pp. 36-44, 2009.
- [2] P. Tambe, "Big data investment, skills, and firm value," *Management Science*, vol. 60, no. 6, pp. 1452-1469, 2014.
- [3] D. Laney, "3D Data Management: Controlling Data Volume, Velocity and Variety." Meta Group. Feb. 6, 2001.
- [4] C. Sniijders, U. Matzat, and U. Reips, "Big Data: Big gaps of knowledge in the field of Internet," *International Journal of Internet Science*. vol. 7, pp. 1 - 5, 2012
- [5] I. A. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, and S. U. Khan, "big data on cloud computing: Review and open research issues," *Information Systems*, vol. 47, pp. 98-115, 2015.
- [6] Brynjolfsson, E. "The productivity paradox of information technology," *Communications of the ACM*, vol. 36, no. 12, pp. 66-77, 1993.
- [7] R. T. Due, "The productivity paradox," *Information Systems Management*, vol. 10, no. 1, pp. 68-71, 1993.
- [8] N. G. Carr, "IT doesn't matter," *Harvard Business Review*, pp. 41-49, May 2003.
- [9] G. George, M. R. Haas, and A. Pentland, "Big data and management," *Academy of Management Journal*, vol. 57, no. 2, pp. 321-326, 2014.
- [10] 진희국, 현근수, 임경빈, 이우현, 김형주, "영화 흥행 실적 예측을 위한 빅데이터 전처리," *정보과학회 컴퓨팅의 실제 논문지*, 제20권, 제12호, pp. 615-622, 2014.
- [11] S. Zhang, C. Zhang, and Q. Yang, "Data preparation for data mining," *Applied Artificial Intelligence*, vol. 17, no. 5-6, pp. 375-381, 2003.
- [12] S. LaValle, E. S. Lesser, M. S. Hopkins, and N. Kruschwitz, "Big data, analytics and the path from insights to value," *MIT Sloan Management Review*, vol. 21, 2013.
- [13] 시로타 마코토, "빅데이터의 충격: 거대한 데이터의 파도가 사업 전략을 바꾼다!" 김성재 역, 한빛미디어, 2013.
- [14] 이수학, 윤병동, "Industry 4.0과 고장예지 및 건전성관리 기술 (PHM) 의 방향," *소음·진동*, 제25권, 제1호, 2015.
- [15] R. Jacobs, "Rise of Robot Factories Leading Fourth Industrial Revolution," *Newsweek*, Mar. 5, 2015.
- [16] S. Ryu, and T. M. Song, "Big data analysis in healthcare," *Healthcare Informatics Research*, vol. 20, no. 4, pp. 247-248, 2014
- [17] A. Carugati, R. Liao, and P. Smith, "Speed-to-fashion: managing global supply chain in Zara," *Proceedings of the IEEE ICMIT*, pp. 1494-1499, Sept. 2008.
- [18] 이진형, "데이터 빅뱅, 빅 데이터(BIG DATA)의 동향," *Journal of Communications & Radio Spectrum*, 2012.
- [19] 한주엽, "삼성·LG 등 국내 대기업 SNS 빅 데이터 분석 한창," *디지털데일리*, 2011. 12. 04.
- [20] 김진욱, "영화 마케팅의 빅데이터 활용효과에 관한 연구," *한국엔터테인먼트산업학회논문지*, 제8권, 제2호, p p. 349-356, 2014.
- [21] 이오준, 박승보, 정다울, 유은순, "소셜 빅데이터를 이용한 영화 흥행 요인 분석," *한국콘텐츠학회논문지*, 제14권, 제10호, pp. 527-538, 2014.
- [22] D. Henschen, "Catalina Marketing Aims For The Cutting Edge Of Big Data," *Newsweek*, Sept. 6, 2011.

저 자 소 개



이충권(정회원)

1995년 계명대학교 경영정보학과 학사 졸업.
1999년 Southeast Missouri State University MBA 졸업
2003년 University of Nebraska-Lincoln 박사 졸업.
2003-2006년 Georgia Southern University 조교수

<주관심분야 : Big Data, Text Mining, IT Jobs>