

빅 데이터의 품질 요소 제안

A propose of Big data quality elements

최상균*, 전순천*

Sang-Kyoon Choi*, Soon-Cheon Jeon**

요 약

빅 데이터가 새로운 가치 창출과 문제 해결의 핵심 엔진이 되는 데이터 중심 시대가 본격적으로 시작되고 있다. 본 논문은 빅 데이터를 활용하기 위하여 빅 데이터의 품질 확보를 위한 품질 요소 정의와 품질 요소별 품질확보 전략에 대하여 논한다. 이를 위해 빅 데이터의 구축 사례, 빅 데이터의 자원 확보 방안 및 빅 데이터의 요소기술, 분석기술과 처리기술 등에 대해 살펴 보았다. 이를 통하여 빅 데이터의 품질 요소를 정의하고 품질 요소별 품질 확보 전략을 제안한다. 빅 데이터의 품질이 확보되면 기업은 대용량의 데이터에서 데이터의 재해석을 통하여 빅 데이터를 추출하고 기업의 경쟁력 제고를 위한 각종 전략을 수립할 것이다.

Abstract

Big data has a key engine of the new value creation and troubleshooting are becoming more data-centric era begins in earnest. This paper takes advantage of the big data, big data in order to secure the quality of the quality elements for ensuring the quality of Justice and quality per-element strategy argue against. To achieve this, big data, case studies, resources of the big data plan and the elements of knowledge, analytical skills and big data processing technology, and more. This defines the quality of big data and quality, quality strategy. The quality of the data is secured by big companies from the large amounts of data through the data reinterpreted in big corporate competitiveness and to extract data for various strategies.

Key words : Big Data, Quality, Data quality

I. 서 론

정보기술과 인터넷이 본격적으로 확산된 지난 20여 년간 기술변화를 통한 우리 사회의 변화속도와 범위는 한마디로 상전벽해 수준이라 할 수 있다. 기술의 변화가 워낙 빠르게 이루어지고 있고 이를 통한 사회변화도 따라가기 힘들 정도로 급속하게 진전되고 있다[1]. 이러한 상황 속에서 거스를 수 없는 또

하나의 메가톤급 변화 물질로서 빅 데이터 시대가 새롭게 전개되고 있다. 최근 2~3년간에 걸쳐 빅 데이터가 가진 무한한 잠재력과 빅 데이터 분석과 활용에 기반한 새로운 가능성에 대해서 다양한 연구 결과들이 쏟아져 나오고 있다. 그런데 빅 데이터에 관한 기술적 논의와 잠재력에 관한 논의는 활발한데 반해 빅 데이터의 품질 확보와 품질 요소 등에 대한 연구는 매우 미흡한 실정이다.

* 김포대학교 e-비즈니스과(e-Business Dept., Kimpo College)

· 제1저자 (First Author) : 최상균(Sang-Kyoon Choi, tel: +82-31-999-4212, email: skchoi@kimpo.ac.kr)

· 접수일자 : 2012년 11월 12일 · 심사(수정)일자 : 2012년 11월 26일 (수정일자 : 2013년 2월 23일) · 게재일자 : 2013년 2월 28일

<http://dx.doi.org/10.12673/jkoni.2013.17.01.009>

이러한 배경 아래 본 연구는 빅 데이터에 대한 품질요소 정의와 품질 확보 방안에 대하여 논하고자 한다. 이를 위해 먼저 빅 데이터의 구축 사례를 통해 기업에 시사하는 바와 빅 데이터의 잠재력을 통하여 정보기술(IT)의 과거, 현재와 미래를 조명하고, 빅 데이터의 자원 확보 방안을 조망하여 빅 데이터의 확보 방안을 제시하였다. 그리고 빅 데이터의 요소기술과 분석기술, 처리기술 등을 살펴보고 빅 데이터의 구축에 필요한 빅 데이터의 품질 요소와 품질 확보 방안을 제안하였다.

본 논문에서의 이러한 논의는 향후 빅 데이터 분석·활용의 활성화, 정부·민간간의 빅 데이터 협력 체계의 공고화 등에 기여할 수 있을 것으로 판단된다.

II. IT와 데이터 패러다임의 변화 및 관련연구

IT는 이제 효율성, 서비스 혁신 등을 추구하는 도구로서의 역할을 넘어서서 개인, 조직 및 사회에 새로운 가치를 창출하고 사회 각 분야와 융합되어 다양한 사회현안을 해결하는데 기여하는 새로운 역할이 요구되고 있다. 사회 패러다임의 변화 및 IT 패러다임의 변화와 함께 데이터 패러다임에 있어서도 근본적인 변화가 일어나고 있다. IT가 주로 효율성 제고 생산성 향상, 프로세스 혁신 및 편리성 제공 등을 위해 활용되던 시대의 경우에는 이들 목적을 위해 구축·운영되는 정보시스템에서의 데이터는 수동적인 관리 대상에 불과한 경우가 많았다. 그러나 정보기술의 급속한 발전으로 데이터의 양적·질적 팽창이 이루어지고 이러한 대용량, 비정형 데이터를 분석하고 처리하는 기술도 병행해서 발전함에 따라 데이터의 가치가 점점 증가하고 있다.

이러한 상황에서 빅 데이터의 등장은 그 의미가 매우 크다고 할 수 있다. 데이터는 그 자체로서 구체적이다. 그리고 빅 데이터의 등장으로 사회현안 해결 및 새로운 가치창출을 위해 데이터 분석기반의 구체적이고 실제적인 접근이 가능해졌다는 점에서 정보화에 대한 새로운 방향 정립 및 대응 전략 모색에 있어 새로운 전환기를 마련할 것이다.

2-1 관련연구

[2]에서는 빅 데이터 분석정보와 기존 데이터베이스의 연관 분석을 통한 비즈니스 가치를 극대화할 필요가 있고, 이전에 관리되지 않던 새로운 데이터를 포함하여 학생 정보를 분석함으로써 예측 능력 및 충분한 정보자료를 활용하여 효율적인 방안을 마련하도록 제안하고 있으나, 빅 데이터의 품질에 대한 언급은 하지 않고 있다. [3], [4]에서는 빅 데이터를 정확하게 이해하기 위한 기본적인 내용들을 설명하고, 빅 데이터가 기업 경영에 가져오는 변화들을 다양한 관점에서 살펴보고, 데이터를 분석하는 기술적인 방법 등에 대하여 논하였으나, 역시 빅 데이터의 품질에 대한 언급은 하지 않고 있다.

III. 빅 데이터 기술

빅 데이터의 품질을 확보하기 위한 방안으로 빅 데이터 잠재력의 분석, 빅 데이터 요소기술, 분석기술 및 처리기술 등을 통하여 빅 데이터 기술을 논한다.

3-1 빅 데이터의 잠재력

매킨지(McKinsey)는 향후에 비즈니스 지형을 바꿀 10대 핵심 기술 트렌드 중의 하나로 빅 데이터를 선정하고 빅 데이터를 수집·저장하고 이를 토대로 새로운 정보를 찾아내는 것이 중요한 가치창출 효과를 가져올 것으로 분석하고 있다[5]. Gartner[6], Deloitte[7], EMC[8] 등도 2012년도 IT 10대 트렌드를 제시하면서 공통적으로 빅 데이터를 향후 핵심 트렌드가 될 것으로 전망하고 있다. 그리고 OECD 발간 보고서에서도 빅 데이터의 경제적 가치와 파급효과를 측정하기 위한 작업이 2011년부터 본격적으로 추진되고 있다[9].

그리고 표 1에서 제시한 것처럼 Economist, Gartner, McKinsey 등에서도 빅 데이터의 잠재력 및 사회경제적인 가치에 대해서 강조하고 있다[10].

표 1. 빅 데이터의 잠재력 및 사회경제적 가치 전망
Table 1. Potential of Big Data and social-economic value of forecast

기관명	주요 전망
Economist	데이터는 자본이나 노동력과 거의 동등한 레벨의 경제적 투입자본, 비즈니스의 새로운 원자재 역할
Gartner	데이터는 21세기 원유로서 데이터가 미래 경쟁 우위를 좌우할 것이며, 기업은 다가올 '데이터 경제시대'를 이해하고 정보 고립(Information Silo)을 경계해야 성공 가능
McKinsey	빅 데이터는 혁신, 경쟁력, 생산성의 핵심 요소로서 의료, 공공행정 등 5대 분야에서 6천억불 이상 가치창출

빅 데이터의 잠재력을 논의하면서 IDC 등 많은 연구들이 빅 데이터의 의미를 데이터의 규모(Big volume)에서보다는 데이터가 가진 가치의 크기(Big value)에 두어야 한다는 점을 강조하고 있다[11]. 초기에는 빅 데이터에 대한 논의가 데이터의 규모와 새로운 분석기술 등에 초점을 두고 시작되었으나 점차 빅 데이터가 지닌 가치와 잠재력 및 활용효과 측면 등으로 무게중심이 이동하고 있다고 할 수 있다. 이와 같은 빅 데이터의 등장에 따라 데이터를 바라보는 시각에 있어서도 패러다임 전환이 일어나고 있다.

3-2 빅 데이터 자원 확보

빅 데이터는 데이터 생성 주체, 데이터의 유형, 저장 방식에 따라 다양하게 분류 가능하며, 생성 주체에 따라 어플리케이션, 센서 등이 생성한 컴퓨터 생산 데이터, 사람이 트위터, 블로그 등에 올린 데이터, 개체간의 관계 데이터 등이 존재하고, 유형에 따라 DB에 저장된 정형, 웹문서와 같은 반정형, 오디오, 비디오 등 비정형 데이터가 존재한다. 또한 저장 방식에 따라 관계형 DB에 저장된 기업 내부 데이터, 관계형 DB에 저장이 어려운 3V(Volume, Velocity, Variety) 데이터, 별도의 포맷으로 저장된 개별 데이터가 존재한다. 이러한 빅 데이터의 특성에 따라 빅 데이터의 수집 및 분석 기술이 활용된다.

표 2. 빅 데이터 자원 분류
Table 2. Classified of Big data resources

	컴퓨터 생산 데이터	사람 생산 데이터	관계 데이터
생성 주체	·어플리케이션 서버 로그(웹사이트, 게임 등) ·센서 데이터(날씨, 물, 스마트 그리드 등) ·이미지, 비디오(트래픽, 보안 카메라 등)	트위터, 블로그, 이메일, 사진, 게시판 글 등	페이스북, 링크드인 등
유형	정형	반정형	비정형
	DB에 저장된 구조적 데이터	웹문서, 메타 데이터, 센서 데이터, 공정 컨트롤 데이터, 쿨상세 데이터 등	소셜 데이터, 문서, 오디오, 비디오, 동영상, 이미지 등
저장 방식	3V 데이터	기업 데이터	이산 데이터
	관계형 DB에 저장하기 어려운 3V 특성을 갖는 데이터	CRM, ERP, DW, MDM 등과 같이 주로 관계형 데이터베이스에 저장된 데이터	스프레드시트, 파일 데이터베이스, 이메일, JSON/XML L 데이터 등 개별적으로 관리되는 데이터

IV. 빅 데이터 요소기술

빅 데이터의 요소 기술은 다음의 세 가지로 구분된다[6].

첫째, 데이터의 크기(Volume)로 데이터의 물리적인 크기 보다는 방대한 양의 크기를 들 수 있다. 웹로그 데이터나 이메일 등의 데이터는 수 PB(PetaByte) 이상이 되지만, 트위터 네트워크 데이터는 수십 GB(GigaByte) 미만이다. 네트워크 데이터는 분석 및 처리가 가장 큰 이슈이다. 따라서 단순한 물리적인 크기가 아닌 데이터의 어떤 속성에 따라 중요성을 판단하고 그것을 처리하는데 어려움이 있느냐 없느냐 인 것이다.

둘째, 데이터의 속도(Velocity)로 데이터를 처리하

는 속도이다. 필요에 따라서 수많은 사용자 요청을 실시간으로 처리한 후 처리 결과를 보내주는 기능도 필요하다.

셋째, 데이터의 형태(Variety)로 전통적인 기업의 데이터 분석은 기업 내부에서 발생하는 운영 데이터인 ERP, SCM, MES(Manufacturing Execution System), CRM 등의 시스템에 저장되어 있는 RDBMS 기반의 정형 데이터였다. 이러한 정형 데이터는 잘 정제되어 있고 의미도 명확하다. 그리고, 스키마를 포함하는 XML, HTML 등의 비정형 데이터도 있다. 하지만 최근에는 이런 데이터뿐만 아니라 기업 외부에서 발생하는 SNS, 블로그, 뉴스, 게시판 등의 데이터나 사용자의 업로드 파일 및 콜 센터의 고객 상담 내용 등의 비정형 데이터도 처리해야 한다.

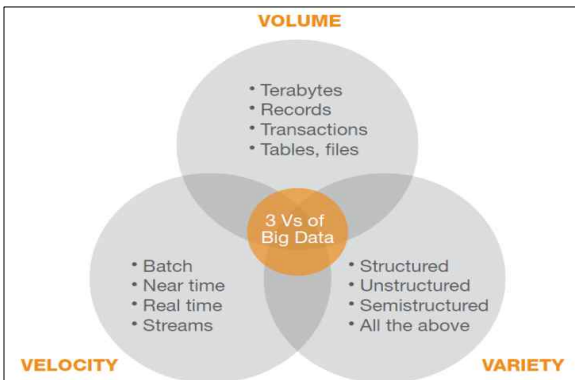


그림 1. 빅 데이터 3대 요소
Fig.1. Three major components of Big Data

4-1 빅 데이터 분석기술

현재의 데이터 분석 기술은 대부분 한 대의 컴퓨터상의 메모리, 파일시스템, 데이터베이스에 데이터를 저장하고 이를 기반으로 데이터를 분석하는 알고리즘을 실행하는 구조다[12]. 대부분의 통계 툴들은 여전히 메모리에 데이터를 로딩해서 통계/분석/마이닝 알고리즘을 실행하는 것이 기본 구조다. 이러한 데이터 분석 시스템 구조는 싱글머신/싱글코어에 최적화 되어 있으며, 최근에는 싱글머신/멀티코어에서 실행할 수 있는 다양한 알고리즘의 개발과 시스템들이 등장하고 상용화되어 쓰고 있다. 지금까지 빅 데이터를 처리하기 위해서는 몇백 기가 메모리와 SAN 스토리지로 대용량의 파일시스템을 마운트할 수 있

는 고 사양 고가의 하이엔드 급 서버를 이용해서 DW, DM을 구축해 왔다. 데이터 증가에 따른 시스템 확장은 더 고사양의 장비로 교체하거나 CPU/메모리/디스크 증설이라는 방식을 이용해서 하는 scale-up 방식만이 유일했다. 문제는 최근 구글, 아마존, 야후, 페이스북, 트위터와 같은 TB(TeraByte)에서 PB(PetaByte) 규모의 데이터를 분석하여 검색 엔진, 소셜 서비스, 광고 등을 하기에는 기존의 시스템, 소프트웨어 아키텍처로는 불가능했던 것이다. 이들이 처리해야 하는 데이터들은 데이터베이스에 깔끔히 정리된 정형화된 데이터가 아니라, 웹을 통해서 수집한 다양한 비정형 데이터와 함께 비디오, 사진, 음향 등 다양한 미디어 정보를 수집해서 분석해야 하기 때문에 더욱 힘들어질 수밖에 없게 된 것이다. 구글은 이러한 문제점을 해결하기 위해 MapReduce라는 구글 파일시스템(GFS) 기술을 활용한다. 이 기술은 Apache Hadoop이라는 기술로 Hadoop은 다수의 서버를 묶어 분산 처리하는 플랫폼이며, 분산 처리하는 'Map' 단계와 결과를 취합하는 'Reduce' 단계로 이루어진 모델로 만들어 졌다. 그림 2는 Hadoop이 빅 데이터를 처리하는 ECOsystem의 구성도를 보이고 있다. Data Store부터 인터페이스까지 6개 세부 파트로 구성된다.

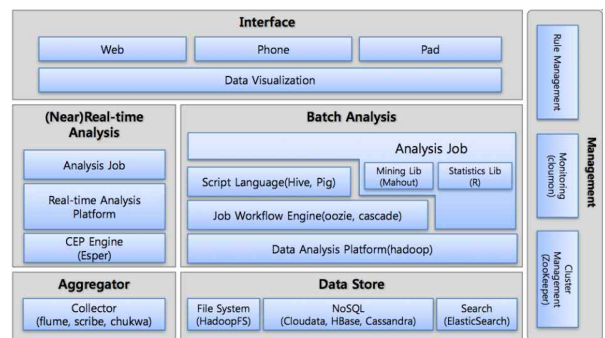


그림 2. Hadoop ECOsystem의 소프트웨어 Stack
Fig. 2. Software stack of Hadoop ECOsystem

빅 데이터 트래픽 폭증에 대응하기 위한 네트워크 기술은 협력통신기술, 반송파집적기술, Femtocell, WiFi Offload, HetNet 등으로 구분되어 활용되고 있다.

4-2 빅 데이터 처리기술

빅 데이터의 처리 기술 가운데 분석 기술이 있다 [12]. 이는 데이터를 분석하는 기술과 방법을 의미하며 통계, 데이터마이닝, 기계학습, 자연어 처리, 패턴 인식 등이 이에 속한다. Hadoop 기술로 대표되는 처리기술은 SQL을 사용하지 않고, 대용량의 데이터를 빠르게 처리하는 기술인 NoSQL을 이용한다. 이러한 기술은 데이터의 안정성이 정합성보다 속도, 비용에 초점을 맞췄기 때문에 일부 데이터 유·손실이 발생해도 처리 결과에 영향을 받지 않는 업무에 적합하다. 로그 데이터 처리 혹은 분석하는 업무를 대표적인 예로 들 수 있다.

빅 데이터 표현 기술은 일반적으로 데이터 시각화로, 분석된 데이터의 특징이나 의미를 쉽게 알 수 있도록 잘 표현해주는 기술이다. 대표적인 시각화 표현기술은 R로 통계 계산 및 시각화를 위한 언어 및 개발환경을 제공하며, 기본적인 통계 기법으로부터 모델링, 최신 데이터 마이닝, 시뮬레이션, 수치해석 기법까지 구현 가능하다. 구현 결과는 그래프 등으로 시각화 할 수 있으며, 다른 프로그래밍 언어와 연결도 용이하다. 계층, 신약연구와 금융예측 분석에 활용되고 있다.

빅 데이터 분석 인프라는 데이터 분석과 표현을 수행할 수 있도록 해주는 기반기술과 플랫폼들이라고 할 수 있으며, 이러한 분석 인프라는 다시 대규모 데이터를 안정적으로 수집해서 저장하는 기술, 저장된 것을 효과적이면서도 빠르게 처리할 수 있는 기술, 저장된 데이터를 다양한 방식과 용도로 사용할 수 있도록 가공하고 관리해주는 기술 등으로 구분되는데, BI(Business Intelligence), DW, 클라우드 컴퓨팅, 분산 데이터베이스(NoSQL), 분산 병렬처리, 분산 파일 시스템 등이 분산 인프라에 속하는 기술들이다.

V. 빅 데이터 품질요소 정의 및 품질 전략

전술한 바와 같이 빅 데이터의 요소기술, 분석기술, 처리기술을 가능하게 하는 빅 데이터의 품질을 확보할 수 있는 품질 요소를 제시하고 이에 대한 품질 확보 방안을 제시한다.

데이터의 품질은 빅 데이터 처리기술에 기반을

두고 이를 활용하는 사용자 관점에서 사용 목적에 적합한 데이터인가에 따라 결정하는 것이 바람직하다. 이를 위한 빅 데이터 특징과 품질관리에 대한 접근법은, 데이터의 대량성, 데이터의 섬세함, 데이터의 소유에 관한 사항 등으로 구분할 수 있다. 빅 데이터의 품질은 데이터 생산자, 데이터 가공자, 데이터 소비자간의 데이터 가치 사슬 단계에서 형성된다. 시스템 및 데이터베이스 등으로부터 수집한 원천 데이터는 모형 개발, 통계적 데이터 처리 과정 등의 분석 과정에서 데이터 품질의 변화 발생 가능하고, 데이터 가공 단계 중 어느 단계에서 품질 문제가 발생했는가에 따라 데이터 신뢰도에 미치는 영향이 결정된다. 이러한 속성에 대한 품질관리 접근방법은 다음의 표 3과 같다.

표 3. 품질관리에 영향을 미치는 빅 데이터 특징과 품질관리 접근방법
Table 3. Big Data Features affecting the quality management and quality management approach

빅 데이터 특징		품질관리 접근방법
대량의 데이터	수작업으로 수집되기 보다는 기계, 프로그램 등에 의해 수집되는 대량의 데이터	→ ·혹시 발생할지 모르는 데이터 사용자의 오류는 무시 ·데이터 수집 과정의 타당성을 방해하는 예외상황을 탐지하는 수준으로 품질 기준 정의(ex. 장치 고장으로 인한 데이터 손실, 장치의 비정상적 상황으로 인한 비정상적 수치 등)
미세하고 정밀한 데이터	클릭 스트림, 미터 값 등 기계, 센서, 프로그램 등에서 생산되는 데이터로 기존 데이터 보다 훨씬 미세한 데이터	→ ·개별 데이터에 대한 타당성 검증은 경우에 따라 불필요 ·개별 레코드에 대한 의미보다 데이터 전체가 나타내는 의미를 중심으로 품질 기준 정의
데이터 소유자 불분명	누가 언제 어디서 데이터를 생산한 것인지에 대한 관리감독이 불가능한 조직 외부의 데이터	→ ·목적이나 통제없이 생산된 데이터에 대한 데이터 품질 기준을 정의하기 위한 다른 방법 필요

이러한 접근 방식을 사용하여 도출된 빅 데이터

사용자 관점에서의 데이터 품질 속성을 제시한다. 즉, 데이터의 내재성, 접근성, 상황성, 표현성 측면에서 데이터 품질 유형을 구분하고, 데이터의 품질을 보장될 수 있는 세부 품질 요소는 표 4와 같다.

표 4. 데이터 품질 유형 및 세부 품질 요소
Table 4. Data quality type and detail quality factor

데이터 품질 유형	데이터 품질 요소
내재적 품질 (Intrinsic)	정확성, 객관성, 진정성 등 데이터 자체의 우수성
접근성 품질 (Accessibility)	접근성, 접근 보안성 등 데이터에 접근할 수 있는 환경적 우수성
상황적 품질 (Contextual)	연관성, 초시간성, 완전성, 데이터양 등 사용자가 목적하는 상황에 적합성
표현적 품질 (Representational)	해석력, 간결성, 용이성, 일관성 등 데이터 표현의 명확성

빅 데이터는 기존의 데이터와 다른 몇 가지 특징으로 인해 데이터 품질 관리 측면에서 다른 방식의 접근이 필요하다. 대량의 데이터, 세밀한 수준의 데이터, 소유자가 불분명한 데이터 특성으로 인해 기존의 품질 기준의 적용은 불가능하고, 모든 개별 데이터에 대한 타당성 보장보다는 빅 데이터 개념 및 특성 측면에서 관리되어야 할 항목과 수준에 대해 품질을 정의한다. 빅 데이터는 충분성 개념의 품질 관리 전략이 필요한데, 빅 데이터 품질관리는 기존 데이터 품질 요소인 정확성(Accuracy), 완전성(Completeness), 적시성(Timeliness), 일관성(Consistency) 측면에서 다른 방식의 품질 전략이 필요하고 이는 정확성, 완전성, 적시성, 일관성에 대한 품질 전략은 데이터의 사용목적, 데이터의 재사용 여부, 일관성 유지 여부에 따라 수립되어야 한다. 이러한 품질 요소별 데이터 품질 전략을 다음과 같이 제시한다.

5-1 정확성

일반적인 데이터베이스의 기본적인 요구사항은 데이터의 정확성이다. 이와 같은 맥락으로 빅 데이터의 기본적인 요구사항은 데이터 정확성이다. 이는 데이터 사용 목적에 따라 데이터 정확성의 기준이

다르게 적용될 수 있으며, 사용자가 접속한 사이트와 이동 지점을 분석하는 클릭 스트림 분석과 부정이나 사기를 탐지하는 경우 데이터의 품질 수준은 다를 수 있다. 데이터 정확성은 선후관계 정확성, 계산/집계 정확성, 최신성, 업무규칙 정확성 등으로 구분될 수 있다.

5-2 완전성

데이터의 완전성은 데이터를 활용하는 사용자 관점에서 매우 중요한 요소라 할 수 있다. 이를 위하여 사용자가 필요한 데이터의 완전한 확보 보다는 필요한 데이터를 식별하는 수준으로 적용 가능하다. 데이터 완전성은 개별 완전성과 조건 완전성으로 구분될 수 있다.

5-3 적시성

데이터의 요구사항 가운데 적시성은 데이터의 존재 여부를 가능하는 요소로 평가된다. 데이터 적시성은 소멸성이 강한 데이터에 대해 어느 정도의 품질 기준을 적용할 것인지 결정하는 요소로, 웹 로그 데이터, 트윗 데이터, 위치 데이터 등의 경우 하루, 몇 시간, 몇 분 동안만 타당성을 갖는 등의 특성이 있다.

5-4 일관성

데이터베이스를 이용하는 사용자 관점에서 데이터의 일관성은 데이터에 대한 신뢰를 보장하는 척도이다. 이는 동일한 데이터라 할지라도 사용 목적에 따라 달라지는 데이터 수집 기준 때문에 데이터 의미가 달라질 수 있기 때문이다. 따라서 데이터 일관성은 빅 데이터가 제공하는 데이터베이스의 기능 가운데에서 매우 의미있는 요소라 할 수 있다. 데이터 일관성은 기준코드 일관성, 참조 무결성, 데이터 흐름 일관성, 컬럼 연관성 등으로 구분할 수 있다.

VI. 결론 및 향후 연구방향

본 연구는 빅 데이터의 품질을 확보할 수 있는 품

질요소와 품질확보 전략을 제시하였다. 빅 데이터 품질은 정확성 보다는 충분성 개념 하에서 조직의 비즈니스 영역 및 목적에 따라 관리하는 것이 바람직하다. 즉, 소유하지 않은 외부 데이터를 활용하고자 할 경우 내부적으로 데이터 신뢰성에 대한 기준 수립 필요하다.

빅 데이터 활용 결과의 정확성 및 신뢰성 향상을 위해 빅 데이터 품질 관리 체계 구축 필요하다. 기업의 정형 데이터, 공공기관이 보유한 공공정보 등은 개별 정보에 대한 품질 관리를 통해 데이터의 중복성, 불일치성 등을 관리할 필요가 있다. 빅 데이터 자원의 품질을 보장하고 활용을 극대화하기 위한 빅 데이터 자원 품질 인증 방안 연구 필요한데, 빅 데이터 품질 관리 가이드라인을 준수한 데이터에 대한 빅 데이터 품질 라이선스와 데이터 공개 가이드라인을 준수한 공개 데이터에 대한 데이터 공유 라이선스 부여 방안 검토 필요할 것으로 판단된다.

감사의 글

본 논문은 2012학년도 김포대학교 연구비 지원에 의함.

Reference

[1] Hyeon-Kon Kim, "New Informatization Strategy Using Big Data", *Telecommunications Review vol 22 no3, SK Telecom*, 2012. 6.

[2] Ji Hyeon Park, Jae Myeong Choi , Byoung Lyoul Park , Heau Jo Kang, "Establishing a Sustainable Future Smart Education System", *The Korea Navigation Institute vol 16 no 3* 2012. 6.

[3] Yu-gun Ham, Sung-byung Chai, "Big Data management change. Data has to know the answer", *SERI*, 2012. 8.

[4] Suzuki, Ryosuke, "Big Data Business", 2012. 3.

[5] McKinsey&Company, "Clouds, big data, and smart assets: Ten tech-enabled business trends to watch", 2010.

[6] Gartner, "Big Data is Only the Beginning of Extreme Information Management", 2011.

[7] Tak-Gil Shim, "Big Data Lifecycle Management", *SK C&C Big data INSIGHT*, 2011.

[8] NIPA, "Ride the wind Big Data, business analytics software market is the fastest growing", *IDC & Info World*, 2012. 6.

[9] MEASURING THE ECONOMICS, "BIG DATA", *Working Party on Indicators for the Information Society, DSTI/ICCP/IIIS, OECD*, 2011.

[10] Ji-sun Jung, "New value creation engine, new possibilities in Big Data and Strategies", *IT & Future Strategy, NIA*, 2011. 12.

[11] Idc, "Worldwide Big Data Technology and Services 2012-2015 Forecast", *IDC #233485*, 2012. 3.

[12] NIPA, "Future society and Big data technologies", *IT & Future Strategy*, 2012. 4.

최 상 균 (Sang-Kyoon Choi)



1993년 8월 : 서강대학교 정보처리학과(이학석사)
 2005년 8월 : 경희대학교 컴퓨터공학과(공학박사)
 1998년 3월~현재 : 김포대학교 경영관광학부 e-비즈니스과 부교수
 1995. 7 전자계산조직응용 기술사
 관심분야 : 데이터 품질, 정보시스템 품질 등

전 순 천 (Soon-Cheon Jeon)



2009. 2~2012. 2 (주)가덕씨엔에스 대표이사
 1996. 8 전자계산조직응용기술사
 2007. 2 서울시립대학교 대학원 경영학 박사수료
 2012년 3월~현재 : 김포대학교 경영관광학부 e-비즈니스과 조교수
 관심분야 : 정보시스템 감리, 소프트웨어 프로젝트 관리, PMO 등