



특집 09

빅데이터 거래 서비스 시스템 구축방안



추병조 (KT)

-
- 목 차 »
1. 서 론
 2. 기존 빅데이터 활용의 문제점
 3. 빅데이터 거래 서비스 시스템
 4. 결 론
-

1. 서 론

최근 정보통신 기술의 일상화가 이루어지면서 하드웨어, 소프트웨어에 이어 데이터가 정보화 사회의 새로운 핵심기술로 등장하고 있다. 이는 과거 하드웨어에서 소프트웨어, 그리고 데이터로 관심영역이 점차 확대되고 있다는 것을 의미한다. 이러한 데이터로의 관심영역의 중심에는 빅데이터가 포함되어 있어서 많은 이들과 영역에서 이슈가 되고 있다¹⁾.

단순한 의미의 빅데이터는 저장되지 않았거나 저장되더라도 분석되지 못하고 버리게 되는 방대한 양의 데이터를 말한다. 실제로 빅데이터가 스마트 사회에서는 매우 많이 발생하고 있는데, 이러한 빅데이터를 분석하여 유의미한 지식을 발생 시킴으로써 가치창출을 이루고 자원을 절감하며 범피해를 낮추고 우리의 미래를 예측하여 대처할 수 있는 수단을 제공할 수 있다는 것은 매우 흥미로운 일이다. 이러한 빅데이터 시대를 맞이하여 다양한 분야에서 빅데이터 관련기술의 개발과 함

께 서비스가 개발되어 기존 비즈니스를 최적화하거나 새로운 비즈니스가 창출되고 있다. 그러나 앞서 얘기한 방대한 양의 데이터가 발생함에도 불구하고 정작 제대로 활용을 못하는 경우가 대부분이며 현재 빅데이터를 활용하는 기업은 대부분 대기업 또는 과거에 데이터를 많이 축적해온 기업만이 빅데이터를 제대로 활용하고 있다. 소규모 기업의 경우 데이터 활용을 위하여 대기업에 데이터를 요청해도 대기업은 이익에 침범 받을 수 있기 때문에 대개 샘플데이터 만을 제공한다. 또한 직접 데이터를 수집하려해도 현실적 제약사항으로 인하여 수집하기가 매우 어렵다. 한국인터넷진흥원의 조사결과에 따르면 정보누출, 유통, 개인정보 및 사생활 침해 등과 같이 개인의 프라이버시의 문제로 인해 데이터를 수집하여 활용하는 것은 매우 어려운 것이 현실이다²⁾.

많은 양의 데이터를 수집하는 것도 문제지만 수집한 데이터를 의미 있게 만들기 위한 정제의 과정 또한 문제이다. 빅데이터 정제 기술 개발은 IT 분야의 뿔이지만 다양한 분야의 서비스 확대

를 위해서는 각 분야의 비즈니스 전문가와 협업이 중요하다. 이미 의료 및 생명공학, 마케팅이나 생산관리 등의 경역학 분야, 공공분야 등 다양한 분야에서 빅데이터의 활용에 관한 관심이 뜨겁게 달아오르고 있지만 IT전문가 혼자서 이러한 분야의 빅데이터를 이해하고 제대로 분석해서 가치 있는 의미를 찾아내는 것은 한계가 있다^[3].

이와 같이 빅데이터를 다양한 분야에 적용하려 하지만 아직까지도 그 활용에는 많은 어려움이 있다. 어려운 궁극적인 원인은 첫째, 저장된 대량의 정형 또는 비정형 데이터에서 의미있는 데이터를 추출하는 과정이 매우 어렵기 때문이다. 특정 시스템 및 비즈니스 분야에서는 빅데이터 솔루션 업체 기반으로 분석기술이 발전하였으나, 아직까지 다양한 분야에서 분석기술은 미흡한 것이 현실이다. 그리고 둘째, 새로운 시스템 및 비즈니스 모델이 계속 생성되는 상황에서 솔루션 업체가 모든 분야의 분석기술을 발전시켜 나간다는 것은 금전적, 시간적 문제로 인해 현실성이 떨어진다. 결국, 빅데이터 분석기술의 다양성 부족으로 인해 빅데이터는 기업 내부에서 자신의 경영전략 및 프로세스 개선을 위해 사용하는데 그치고 있다.

본 고에서는 이러한 문제점을 해결하기 위해,

대량의 데이터를 수집하고 유의미한 데이터를 추출하여 정제하는 방법과 정제된 데이터 거래가 안전하게 이루어 질 수 있는 빅데이터 거래 서비스 시스템의 구축방안에 대해서 살펴보도록 한다. 먼저 제 2장에서는 기존 빅데이터 활용의 문제점에 대하여 더 자세히 알아보고, 3장에서는 빅데이터 거래 서비스 시스템에 대하여 설명하고 4장에서 결론을 맺는다.

2. 기존 빅데이터 활용의 문제점

우선 빅데이터를 기반으로 사업진출 하고자하는 사업가를 예로 들어서 기존 빅데이터 활용의 문제점에 대하여 알아보도록 하겠다. 사업가 OO씨가 빅데이터를 기반으로 사업을 진출하고자 한다면, 사업 아이템에 관련된 대량의 데이터를 수집할 필요성이 있다. 여기서 데이터 수집 방법에는 <표 1>과 같은 방법들이 존재하며 각 방법은 그에 따른 문제점이 발생하게 된다.

데이터 수집이 완료되면 데이터 정제의 과정을 거치게 되는데 빅데이터 정제과정은 빅데이터 분석 기술과 마찬가지로 매우 중요하며 중복 데이터 검출 및 최적화 기술이 그에 해당한다^[5]. 데이터를 수집한 OO씨는 아이템의 비즈니스 모델에

〈표 1〉 데이터 수집방법의 문제점

구분	내용
개인/기업 축적 데이터 활용	개인정보유출로 인한 개인 데이터 수집의 어려움
	기업은 대개 샘플데이터만 제공
설문조사	모집단의 모든 응답자에게서 측정되지 않음, 측정범위가 제한적 ^[4]
	설문조사 과정 중 많은 시간을 소비
	설문조사 과정 중 금전적 손해 발생
	단순 기억에만 의존하기 때문에 정확한 데이터를 얻기 힘들
인터넷 검색	수집과정을 반복해야 할 우려 존재
	정확한 데이터를 얻기 힘들
	의미 있는 데이터 정제 과정 필요

〈표 2〉 데이터 정제방법의 문제점

구분	내용
비즈니스 전문가	다른 아이템을 기획할 경우 또 다른 비즈니스의 전문가를 고용
	아이템과 비즈니스 전문가 1:1 비율 그에 따른 전문가 고용 비용 증가
데이터 분석(정제) 시스템 개발	아이템과 비즈니스 모델의 1:1 비율에 따른 데이터 분석(정제) 시스템 개발 비용 증가
	데이터 분석 시스템 개발에 따른 시간 소비

대한 도메인 전문가를 고용하여 데이터 분석 작업을 거친 후 의미 있는 데이터로 정제 과정을 거칠 것이고, 정제된 의미 있는 데이터를 기반으로 사업을 확장한다. 하지만 여기서 데이터 정제에는 <표 2>와 같은 문제점이 발생하게 된다.

위와 같이 데이터 수집과 정제에는 많은 문제점이 발생한다. 다음 3장에서는 데이터 수집과 정제의 문제점에 대한 해결방안과 빅데이터 거래 서비스 시스템의 구조에 대하여 살펴보도록 한다.

3. 빅데이터 거래 서비스 시스템

먼저 본 고에서 논의하는 빅데이터는 일반적인 사용자가 개인용 또는 업무용 프로그램을 사용하면서 자동으로 축적되는 데이터를 대상으로 한다. 단순 구두 또는 수기로 작성되어 취합되는 데이터는 제외한다.

기존의 빅데이터 활용의 문제점은 크게 3가지로 구분할 수 있다. 첫째 데이터를 가지고 있는 개인 또는 기업은 자신의 데이터를 제공하려 하지 않는다. 특히 실시간으로 데이터를 수집하는 것에 대해 매우 부정적이다. 둘째 데이터 분석 및 정제를 위해서는 비즈니스별 도메인 전문가와 시스템이 필요하다. 셋째 데이터를 활용하고자 사람은 원하는 데이터를 누가 가지고 있는지 파악하기가 매우 어렵다.

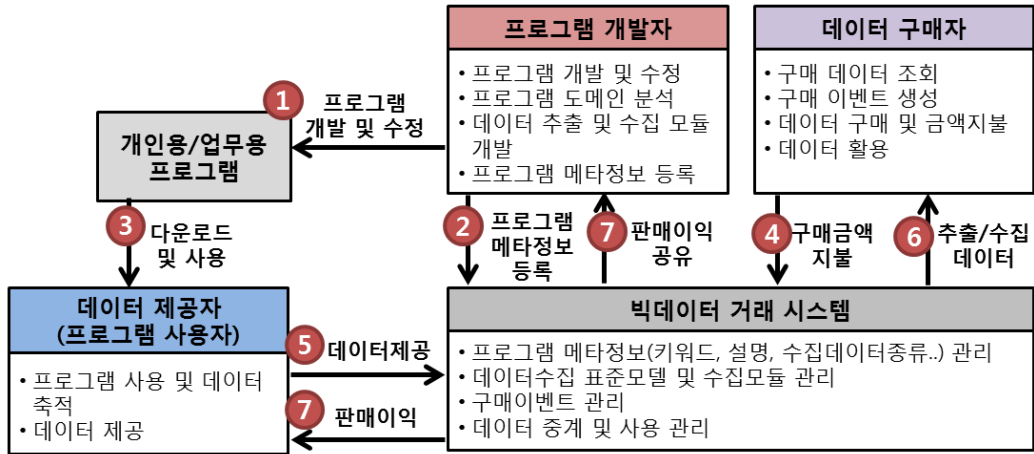
기존의 빅데이터 활용의 문제점을 해결하기 위해 본 고에서는 빅데이터 거래 비즈니스 모델과 빅

데이터 거래 시스템의 결합으로 해결하고자 한다.

3.1 빅데이터 거래 비즈니스 모델

(그림 1)은 빅데이터 거래 비즈니스 모델을 나타낸다. 빅데이터 거래 비즈니스 모델의 핵심은 데이터 판매이익을 데이터제공 및 정제한 관련자에게 직접 제공함으로써 데이터제공 및 데이터 정제를 활성화 하는 것이다.

앞 써 언급한 3가지 문제점은 다음과 같이 해결한다. 첫 번째 문제인 데이터 제공에 대한 문제는 데이터 제공자에게 직접 판매이익을 제공함으로써 해결한다. 데이터 제공자에게 데이터 제공으로 인한 판매이익을 제시하고, 이를 기반으로 데이터제공 동의를 받는다. 두 번째 문제인 도메인 분석 및 정제는 개발자가 직접 함으로써 해결한다. 프로그램 개발자가 해당 도메인에 대해서는 가장 많은 지식을 가지고 있기에 분석 및 정제 작업을 하기에 매우 적합하다. 또한 개발자는 사용자가 데이터제공 동의를 할 경우 데이터 정제 및 수집이 자동으로 되도록 프로그램을 수정한다. 세 번째 문제는 개발자가 프로그램에 대한 메타데이터를 빅데이터 거래 시스템에 등록함으로써 해결한다. 여기서 말하는 메타데이터는 프로그램 설명, 키워드, 수집데이터 종류, 정제데이터 종류 등을 말한다. 메타데이터는 향후 구매자가 원하는 데이터를 검색하거나 데이터를 구매하기 위한 기본정보로 활용된다.



(그림 1) 빅데이터 거래 비즈니스 모델 흐름도

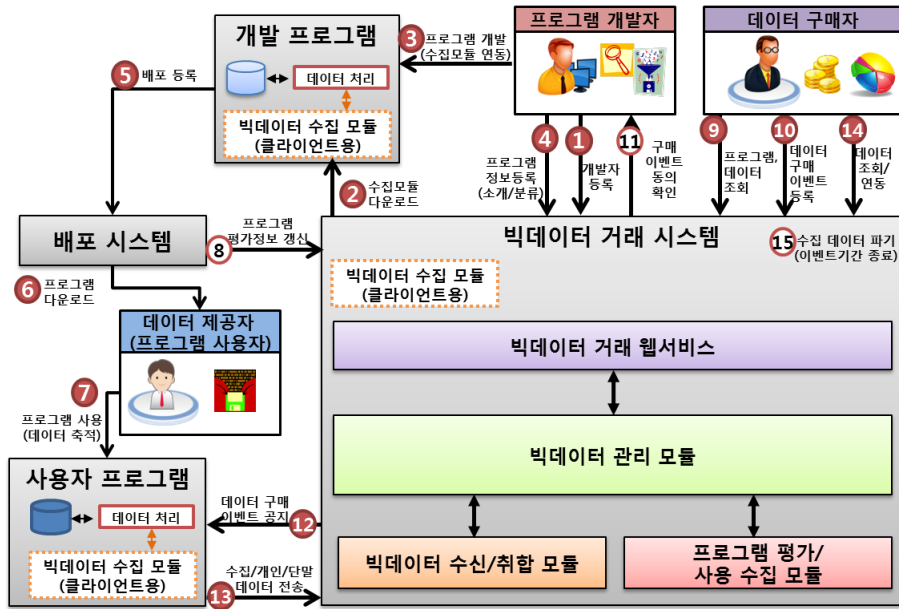
빅데이터 거래 비즈니스 모델에서는 사용자 동의 전 데이터 수집은 원칙적으로 허용하지 않는다. 기존의 빅데이터 활용방법인 실시간 데이터 수집 후 다양한 도메인 전문가를 통한 정제 및 활용은 데이터수집 동의 및 기타 많은 문제를 발생 시킨다.

본 고에서 제시하는 모델은 실시간으로 데이터를 수집하지 않고, 사용자는 자신의 필요에 의해 프로그램을 사용하고 내부적으로 데이터를 축적한다. 이때까지는 프로그램 사용자는 데이터 구매자 및 빅데이터 거래 시스템의 존재를 모르고 프로그램을 사용한다. 데이터 구매자가 데이터 구매 요청을 할 경우 이때 데이터 제공 동의 이벤트가 사용자에게 전달된다. 만약 사용자 데이터 제공에 동의할 경우 장기간 축적된 데이터가 빅데이터 거래 시스템에 일괄 자동 전송한다. 만약 구매자가 3개월 이상 사용한 육아프로그램의 데이터를 사용자당 10,000원에 구매한다면 구매자가 이벤트를 생성한 시점에 데이터 제공 동의 메시지가 사용자에게 전달된다. 이때 데이터 제공 시 지불되는 금액이 사용자 화면에 표시된다. 사용자가 동의할 경우 내부적으로 축적된 3개월 이

상의 데이터가 빅데이터 거래 시스템으로 전달된다. 구매대금은 사용자와 개발자에게 공동 지불된다. 예를 들어, 80%는 데이터를 제공한 프로그램 사용자에게 지불되고, 20%는 도메인 분석 및 정제 모듈을 개발한 프로그램 개발자에게 지불된다.

3.2 빅데이터 거래 서비스 시스템 구조

기존 빅데이터 활용의 문제점을 해결하기 위한 빅데이터 거래 서비스 시스템의 구조 및 서비스 흐름은 (그림 2)와 같다. 빅데이터 거래 서비스의 구성원으로 데이터를 구매하고자 하는 데이터 구매자, 데이터를 제공하는 데이터 제공자, 데이터를 수집하기 위한 프로그램을 개발하는 프로그램 개발자로 나뉜다. 개발 프로그램은 개발자 환경에서 개발 중인 프로그램을 말하며, 사용자 프로그램은 배포시스템을 통해 배포 후 사용자가 직접 사용하고 있는 프로그램을 말한다. 프로그램은 신규 프로그램뿐만 아니라 기존의 운영 중인 프로그램에 빅데이터 거래 시스템을 연동하기 위해 업데이트한 프로그램을 모두 포함한다. 아래 <표 3>에서 각 구성원의 역할과 (그림 2)의 서비



(그림 2) 빅데이터 거래 서비스 시스템 구조 및 서비스 흐름

스 흐름별 내용을 확인할 수 있다.

일반적인 예시를 통해 서비스 흐름을 설명하면 다음과 같다. 예시가 되는 프로그램은 육아 관리 프로그램으로 가정한다. 개발자는 클라이언트용 빅데이터 수집 모듈을 빅데이터 거래 시스템으로부터 다운로드 받아 이를 활용하여 프로그램을 수정한다. 육아 관리 프로그램의 데이터 처리 모듈을 수정하고, 빅데이터 수집 모듈과 연동하여 육아 관리 프로그램을 업데이트 한다. 이 때 빅데이터 수집 모듈의 API를 활용하여 수집되는 데이터의 종류를 서버에 등록한다. 예를 들어 몸무게, 키, 수면 및 배변 정보를 연동 방법을 정의하고, 수집되는 데이터를 서버의 분류 기준에 맞추어 등록한다. 이때 등록한 데이터는 향후 데이터 구매자 검색의 기준이 된다. 개발자는 업데이트된 육아 관리 프로그램 배포 시스템에 등록한다. 배포 시스템은 안드로이드 마켓, 애플 앱스토어 및 사설 서버 등으로 구성될 수 있다. 사용자는 배포 시스템으로부터 육아 관리 프로그램을 다운받아

사용한다. 사용자는 육아 관리 프로그램을 이용하여 몸무게, 키, 수면, 배변 등과 같은 다양한 정보를 저장 및 관리한다. 이때 육아 관련 정보가 프로그램 내부에 축적이 된다. 그리고 필요에 따라 배포 시스템에 의해서 프로그램 평가정보가 빅데이터 거래 시스템으로 전송 될 수 있다. 육아 관련 정보가 필요한 데이터 구매자는 빅데이터 거래 시스템을 통해 원하는 데이터를 검색한다. 개발자가 분류기준 맞춰 등록해 놓았던 육아 관련 프로그램을 검색하고, 프로그램 평가정보 및 사용정보를 참고하여 데이터 구매 이벤트를 등록한다. 빅데이터 거래 시스템에 구매 이벤트가 등록되면 관련 개발자에게 이벤트가 공지된다. 개발자가 동의할 경우 구매 이벤트는 육아 관리 프로그램을 통해 사용자에게 공지된다. 개발자가 동의 조건을 무료 또는 특정 기준치를 정해 놓았을 경우 개발자 동의는 생략될 수도 있다. 구매 이벤트 공지에는 사용자 데이터 구매 조건, 데이터 사용 기간, 데이터 사용 목적, 개인정보 제공

〈표 3〉 빅데이터 거래 서비스 시스템 구성원의 역할 및 서비스 흐름

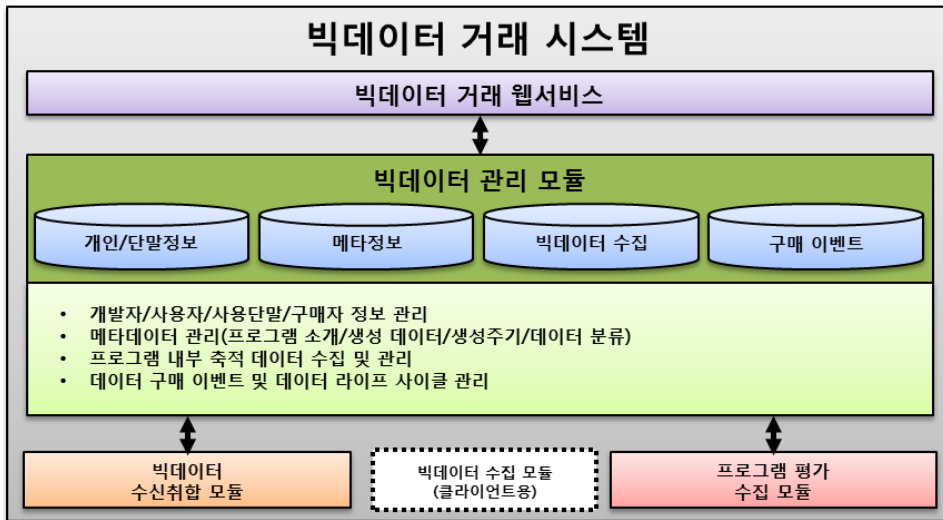
구성원	역할	흐름
프로그램 개발자	개발자 등록	①
	빅데이터 거래 서비스 시스템에서 표준화된 수집모듈을 다운로드	②
	기존 개발된 프로그램 혹은 개발할 프로그램에 수집모듈과 연동하여 향후 사용자 데이터 제공 동의를 하였을 경우 빅데이터 거래 시스템에 정제된 데이터가 일괄 전송될 수 있도록 프로그램 수정	③
	프로그램 메타정보 등록(프로그램 소개, 키워드, 수집되는 데이터 종류, 수집 주기 등등)	④
	배포시스템에 개발 프로그램 등록	⑤
	데이터 구매 이벤트 확인 후 수락 및 거절함. 미리 정의된 동의 조건이 만족할 경우 생략될 수 있음	⑪
데이터 제공자 (프로그램 사용자)	배포 시스템에서 사용할 프로그램 다운로드	⑥
	프로그램 사용. 프로그램 사용에 따른 사용 데이터 축적	⑦
	데이터 구매 이벤트 확인 후 데이터 제공 동의 또는 거절	⑫
	장기간 수집된 빅데이터 일괄 전송	⑬
데이터 구매자	빅데이터 거래 웹서비스를 통한 프로그램 및 데이터 조회	⑨
	빅데이터 거래 웹서비스를 통한 데이터 구매 이벤트 등록	⑩
	데이터 구매에 따른 금액 지불	⑭

여부 등과 같은 다양한 정보가 포함될 수 있다. 사용자가 구매 이벤트에 동의할 경우 육아 관리 프로그램 내부 저장 모듈에 저장된 데이터는 빅데이터 수집 모듈을 통해 빅데이터 거래 시스템으로 전송 된다. 전송된 데이터는 빅데이터 수신 취합 모듈에 의해 수집되고, 빅데이터 관리 모듈을 통해 저장 된다. 데이터 구매자는 빅데이터 거래 웹서비스를 이용하여 육아 관련 정보를 조회 및 활용한다. 구매 이벤트에서 약속한 데이터 사용 기간이 종료될 경우 빅데이터 거래 시스템은 수집한 육아 관련 정보를 서버에서 삭제한다.

3.3 빅데이터 거래 시스템의 프로그램 모듈

빅데이터 거래 시스템에는 크게 빅데이터 관리 모듈, 클라이언트용 빅데이터 수집 모듈, 빅데이

터 수신취합 모듈, 프로그램 평가 수집 모듈, 빅데이터 거래 웹서비스로 구성된다. 빅데이터 관리 모듈은 사용자에게 의해 수집되는 데이터들의 코드를 정의하고 분류 및 저장하는 역할을 한다. 예를 들어 수유량은 "100900", 배변량은 "100901", 수면량은 "100902" 등과 같은 수집데이터 코드를 관리한다. 또한 수집하고자 하는 메타데이터 분류에 따라 의미 있는 데이터가 추출될 수 있도록 개발자가 항목, 분류방식 등을 정의할 수 있도록 한다. 메타데이터는 "데이터에 대한 데이터" 혹은 "정보에 대한 정보"라는 포괄적인 정의로부터 시작하여 데이터베이스에 수록된 각각의 데이터에 대한 기술을 의미한다⁶⁾. 본 고에서 메타데이터는 데이터 구매자가 구매여부 판단에 필요한 각종 정보를 포함한다. 그리고 클라이언트용 빅데이터 수집 모듈은 표준화된 인터페이스로 빅데이터 거래 시스템 연동하기 위한 것으로서, 개발



(그림 3) 빅데이터 거래 시스템 주요 프로그램 모듈

자가 개발한 프로그램에 포함되어 사용자에게 의해 축적되는 데이터가 빅데이터 거래 시스템에 연동되어 빅데이터가 수집 될 수 있도록 한다. 프로그램 평가 수집 모듈은 프로그램 사용자의 평가 및 데이터 구매자의 평가 정보를 실시간 업데이트 한다. 이를 통해 빅데이터 거래 시스템에 등록된 프로그램들의 수집데이터 질을 확인할 수 있다. 빅데이터 거래 웹서비스를 통해 데이터 구매자는 프로그램 및 수집데이터 관련 정보를 검색한다. 또한 구매 이벤트 조건에 의해 데이터 외부 출력이 가능한 경우 빅데이터 거래 웹서비스를 통해 데이터 다운로드가 가능하다.

4. 결론

본 고에서는 빅데이터의 수집과 정제 방법 대하여 정의 하였다. 그리고 정제된 빅데이터를 거래하기위한 검색과 거래방법, 데이터 수집 모듈에 대하여 설계하였다. 본 고에서 제안한 빅데이터 거래 서비스 시스템을 활용 할 경우 서비스 진

출 시간과 비용을 대폭 감소시키는 것이 가능할 것으로 기대된다.

스마트 사회에서 빅데이터는 필수불가결하게 발생하는 요소이며, 이것을 어떻게 잘 활용하는가 스마트사회에서 시장진출의 성패를 좌우한다고 해도 과언이 아니다. 오늘날에도 수많은 데이터들이 쌓이고 버려지는데 그 버려지는 데이터마저도 분명히 원하는 사람이 존재할 것이고 정제를 통하여 유의미한 데이터로 창출 가능 할 것이다. 따라서 이제는 버려지는 데이터가 하나도 없이 다양한 가치창출을 이루는 정보화 사회를 이루어나야 할 때가 도래한 것이다.

빅데이터 거래 서비스 시스템은 쌓이는 데이터 뿐만 아니라 버려지는 데이터 또한 새로운 가치 창출을 도와주며, 더 나아가 IoE 시대에 발 맞추어 사용자 프로그램뿐만 아니라 사물데이터 또한 관리 하여 사물데이터의 수집 및 거래가 가능하도록 하는 것이 필요하다.

참고 문헌

- [1] 조영임, "빅데이터의 이해와 주요 이슈들", 한국 지역정보학회지, 제 16권, 제 3호, pp.43-65, 2013.
- [2] 이환수, 임동원, 조항정, "빅데이터 시대의 개인 정보 과잉이 사용자 저항에 미치는 영향", 지능정보연구, 제 19권, 제 1호, pp.125-139, 2013
- [3] 조완섭, "빅데이터 활용과 데이터 과학자", 정보과학회지, 제 32권, 제1호, pp.59-65, 2014.
- [4] 김광용, 김기수, "인터넷 설문조사의 방법론적인 문제점과 데이터마이닝 기법을 활용한 개인화된 인터넷설문조사 시스템의 구축", 품질경영학회지, 제 32권, 제 2호, pp.93-108, 2004
- [5] 박호진, 권영현, 안영민, "빅데이터와 빅데이터 정제 기술", 한국컴퓨터정보학회지, 제 21권, 제 1호, pp.1-8, 2013.
- [6] 남태우, 이승민, "메타데이터의 의미론적 확장에 관한 연구", 한국문헌정보학회지, 제 44권, 제 4호, pp.373-393, 2010.

저자 약력



추 병 조

이메일 : cbj38317@naver.com

- 2007년 동서대학교 컴퓨터정보공학부(학사)
- 2009년 부산대학교 컴퓨터공학과(석사)
- 2009년~2013년 현대유엔아이(주) 연구소 선임연구원
- 2013년~현재 (주)KT G&E부문 소프트웨어개발센터 매니저
- 관심분야: IOT/IOE, LBS, 빅데이터, 플랫폼 비즈니스