

<http://dx.doi.org/10.7236/JIWIT.2012.12.6.267>

JIWIT 2012-6-33

빅데이터를 이용한 SNS 활용방안 연구

SNS using Big Data Utilization Research.

신승중*

Seung-Jung Shin

요약 IT 융합, 소셜 미디어, 서비스 산업 고도화, 기업들의 고객 데이터 수집활동, 멀티미디어 콘텐츠의 폭발적 증가와 스마트폰 보급, SNS 활성화, 사물통신망의 저변확대로 데이터량은 10년 전 산업분야에 걸쳐 고르게 EDW (Enterprise Data Warehouse)의 수요가 증가하였다. 특히 통신업계에서는 KT가 전사적인 EDW를 진행 했고, 산자부도 각각의 업무 부서별로 여러 건의 DW 프로젝트가 진행 하였다. 이외에도 연세의료원, 건국대 병원 등 많은 종합 병원들이 올해 DW의 도입하여 구축을 완료하였다. 그러나 계속 증가되고 있는 데이터와 사용자의 증가는 데이터의 관리에 또 다른 문제점을 만들고 있다. 최근 SNS사용자의 급증과 이를 배경으로 한 응용 연구들이 진행되면서 빅데이터를 이용한 새로운 연구를 제안하고자 한다.

Abstract IT convergence, social media, and the companies' customer service industry advancement, data collection activities, explosion of multimedia content with increased smartphone penetration, SNS activation networks to expand the pool of things, 10 years ago, the amount of data eunneun evenly across industries, EDW (Enterprisehad increased the demand for the Data Warehouse).Recent proliferation of SNS users and applied research background with Big Data as a new study is proposed to proceed.

Key Words : Big data, SNS, MapReduce, Smart phone

1. 서론

오늘날 맥킨지는 비즈니스 지형을 바꿀 10가지 기술 트렌드 중 하나를 Big Data로 정하였다. 이는 Big Data를 수집하고, 저장하는 기반을 통하여 새로운 정보를 찾아내는 것이 경제성장을 위한 중요한 가치창출 효과를 가져 온다는 분석을 하고 있다. 이러한 환경과 소셜은 또 하나의 관계를 설정할 수 있다는 것이다. 스마트폰의 보급으로 다양한 종류의 스마트폰 어플리케이션이 개발되고 있는데, 글쓰기의 주요 관심사항이나 글쓰기의 친한

친구관계를 알 수 있으며, 특히 인기 또는 급상승 키워드를 알 수도 있다.

그 뿐아니라 링크종류별 그림이나, 뉴스 그리고 동영상의 순위도 알 수 있으며, 키워드 나 자주 방문하는 지역 등을 알 수 있다는 것이다.

또한 소셜 네트워크 데이터로 글 퍼짐 관계와 특정 주제에 대해 관심있는 그룹, 그리고 그룹간의 관계와 그룹 내에 영향력 있는 사람 관련 분야에 대한 전반적인 내용을 분석 할 수 있다^[1]. 이런 내용은 문서분류, 문서군집, 정보추출, 문서요약 등과 특정 서비스 및 상품에 대한 시

*정회원, 한세대학교 IT학부

접수일자 : 2012년 10월 25일, 수정완료 : 2012년 11월 24일
게재확정일자 : 2012년 12월 14일

Received: 25 October 2012 / Revised: 24 November 2012 /

Accepted: Revised: 14 December 2012

*Corresponding Author: expersin@hansei.ac.kr

Dept of IT, HanSei University, Korea

장규모 예측, 소비자의 반응, 입소문 분석등에 활용하는 데 있으며, 특히 입소문의 중심이나 허브 역할을 하는 사용자를 찾는데 활용되고 있다.

그리고 최근에 기상정보서비스를 제공하기 위한 기술로 플랫폼 관련기술을 소개하고자 한다.

Facebook: hadoop/hbase, ptail, puma

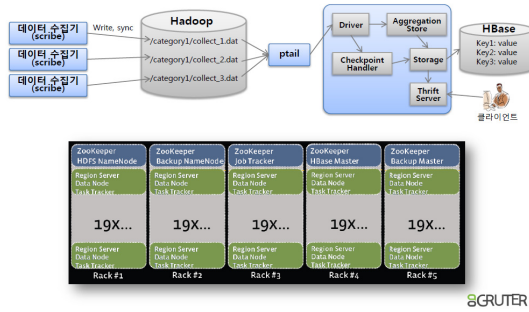


그림 1. 전체 시스템 구성도
Fig 1. The overall system configuration

표 1. BigData 분석기술
Table 1. BigData Analytical techniques

구분	설명	오픈 소스
데이터 수집	데이터 발생원으로부터 안정적인 저장소로 저장하는 기능 수행	Flume, Scribe, Chukwa
원본 데이터 저장	수집된 데이터를 안정적으로 저장하는 저장소 비구조적 데이터 저장소로 주로 대용량 파일 저장소	Hadoop FileSystem, MogileFS
트랜잭션 데이터 저장	원본 데이터를 실시간으로 저장, 조회 처리를 하기 위한 저장소	NoSQL(Cloudata, HBase, Cassandra), Katta, ElasticSearch
실시간 분석 플랫폼	구조적 저장소 또는 검색 엔진 기술을 활용	S4, Storm
전체 또는 부분 데이터에 대해 복잡하고 다양한 분석 수행	대용량 처리를 위해 분산, 병렬처리가 필요 단순 텍스트 분석부터 그래프 분석까지 다양한 분석 모델 지원	Hadoop MapReduce(Hive, Pig), Graph, GoldenOrb
데이터 마이닝/통계 도구	Cluster, Classification 등과 같이 데이터 마이닝을 위한 기본 알고리즘 라이브러리 및 도구	Mahout, R
클러스터 관리 및 모니터링	대부분 분산 시스템으로 구성되기 때문에 전체 클러스터에 대한 관리 및 모니터링도 복잡	ZooKeeper, HUE, Cloumon
데이터 Serialization	이기종 플랫폼 및 다양한 종류의 솔루션을 사용하기 때문에 데이터 전송 및 처리에 대한 표준 프레임워크 필요	Thrift, Avro, ProtoBuf

II. 연구활용 기술

본 연구배경 기술로는 표 2 에서 제시하고 있는 방대한 텍스트에서 의미 있는 정보를 추출하여 다른 정보와의 연계성 파악하는 텍스트 마이닝 기술과 데이터의 긍정, 부정, 중립의 선호도를 판별하는 오피니언 마이닝기술 그리고 소셜 네트워크 연결구조 및 연결강도등을 바탕으로 사용자의 명성 및 영향력을 측정하여 분석하는 소셜 네트워크 분석기술이 있다. 특히 페이스 북이나 트위터에서 사용 될 수 있는 즉 관측 개체를 몇 개의 그룹으로 나눔으로써 대상집단에 대한 효율적인 활용을 가능케 하는 기술로는 군집분석기술 등이 있다.

1. 플랫폼 관련기술



* 출처 : NICT(위), Viewdle(아래)
그림 2. NICT와 뷰들의 신규 서비스 화면 스냅샷
Fig 2. NICT and the new service screen snapshot of views

구글에게 MapReduce가 없었다면 현재의 구글은 없었을지도 모른다. 구글이 대단한 것은 세계에서 가장 많고 자부할 수 있는 그 방대한 데이터를 유지하고 있는 것은 MapReduce가 있기 때문이다.

표 2. BigData 플랫폼 관련기술
Table 2. BigData Platform technologies

구분	주요 내용
텍스트 마이닝 (Text Mining)	문서 및 인터넷 페이지 등에 존재하는 비정형 텍스트 데이터에서 자연 언어 처리* 기술과 문서처리 기술을 조종해 유용한 정보를 추출, 가공하는 기술 * 자연 언어 처리(Natural Language Processing) : 인간의 언어를 기계적으로 분석해서 컴퓨터가 이해할 수 있는 형태로 표현하는 제반 기술로 인공지능과 밀접한 관련이 있음
오피니언 마이닝 (Opinion Mining)	정형/비정형 텍스트 상의 긍정, 부정, 중립과 같은 선호도를 판별하는 기술
소셜 네트워크 분석 (Social Network Analysis)	SNS 연결 구조 및 연결 강도 등을 바탕으로 사용자의 명성 및 영향력을 측정하는 기술
군집 분석 (Cluster Analysis)	비슷한 특성을 가진 개체를 통합하면서 최종적으로 유사한 특성의 군을 발굴하는 기술

* 출처 : Wikipedia, 스트라티비스 재구성

표 2 에서 제시되고 있는 특징들을 살펴보면, 텍스트 마이닝은 구조화되지 않은 자연어로 이루어진 텍스트에서 의미를 찾아내며, 평판분석(오피니언 마이닝)은 전문가에 의한 선호도를 나타내는 표현/단어 자원의 축적이 필요하다. 또한 소셜 네트워크 분석의 특징은 네트워크 내 조밀한 연결 구역을 명명하는 작업을 한다. 그리고 군집분석의 특징은 비슷한 특성을 가진 개체를 합쳐가면서 유사군을 발굴하는데 있다는 것이다. 이를 기반 기술을 응용한 구글의 MapReduce 그림 3과 같다.

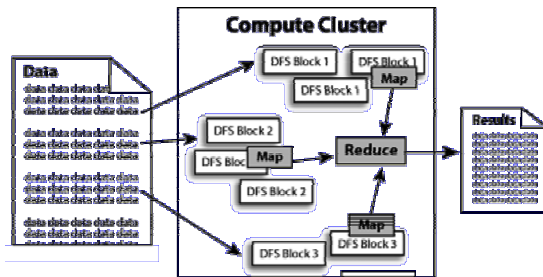
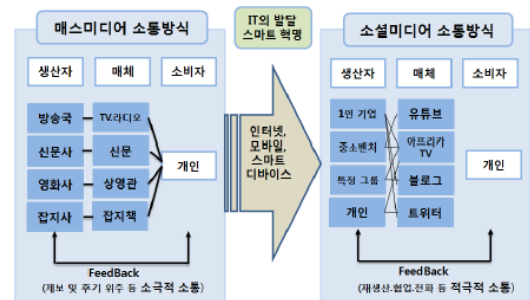


그림 3. 구글의 MapReduce
Fig 3. MapReduc of Google



*출처: 한국정보화진흥원

그림 4. IT의 발달과 소통방식의 변화
Fig 4. IT development and the changes of communication

그림 4는 빅 데이터 분석은 기업의 가치 극대화를 위해 활용될 수 있는 만큼 기존 환경과의 연계, 고가용성, 사용의 용이성, 보안, 시각화 및 적시분석 등이 가능해야 한다. 이미 전통적인 분석 기술로 많은 성과를 거뒀지만 빅 데이터 방식의 분석을 통해 더 많은 성과를 거둬야 한다는 압박감은 증가하고 있다. 분석기술과 관리 능력에 초점을 맞춰 심층적인 개선에 주력해야 할 것이다.

본 연구에서는 빅 데이터의 등장과 더불어 효율적인 활용으로 새로운 비즈니스 창출, 기업 경쟁력 극대화 등

을 이룰 수 있는 빅 데이터 관련기술을 응용하여 새로운 모델을 제시하고자 한다.

2. 빅데이터 도입

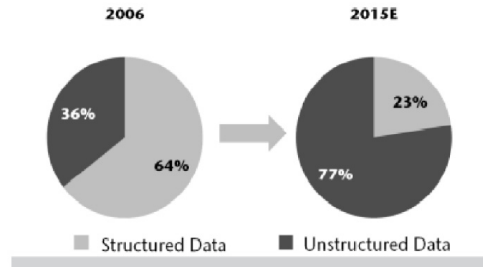


그림 5. 정형/비정형 비중 변화추이(2006 vs 2015)
Fig 5. The changes of Orthopedic/Atypical proportion

미국의 투자은행 제퍼리스(Jefferies)가 최근 IT업계를 뜨겁게 달구고 있는 빅데이터에 대한 관심이 크게 증가하고 있다는 내용의 심층 분석보고서를 발간하고 있으며, 특히 그림 5에서 보는 바와 같이 기업내부에 축적된 비정형데이터가 2006년에 36%에서 2015에는 77%인 2배 이상 증가될 예상이므로, 매년 전 세계는 데이터의 축적이 40%이상증가하고 있는 실정이다. 그리고 최근 축적되고 있는 데이터는 95%이상이 특정문형이나 수치로 정형화 시킬 수 없어 기존 데이터베이스 체계로 해석하기 힘든 것들인 반면, 데이터 속성 측면에서는 고객의 실생활과 직접적인 연관성을 띠는 경우가 많아 오히려 활용도는 더욱 높아지는 추세이다^[3].

III. 빅데이터의 패러다임 변화

1. 빅데이터의 추이단계 및 시장규모

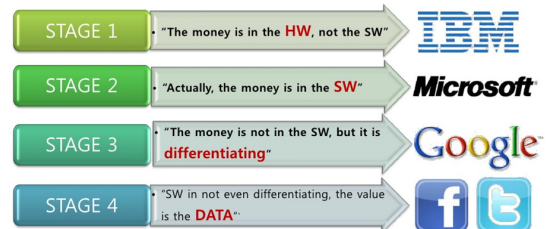


그림 6. 소프트웨어 공급자 4단계(출처: ETRI)
Fig 6. Software Provider in step 4

그림 6 에서 보여주는 바와 같이 가치의 흐름이 H/W 에서 S/W로 그리고 차별화에서 DATA로 이진되고 있을 을 알 수 있다. 물론 기업의 생존변화가 제품의 새로운 가치를 부여하듯이 새로운 변화를 요구 하고 있다.

2. 빅데이터의 시장 규모

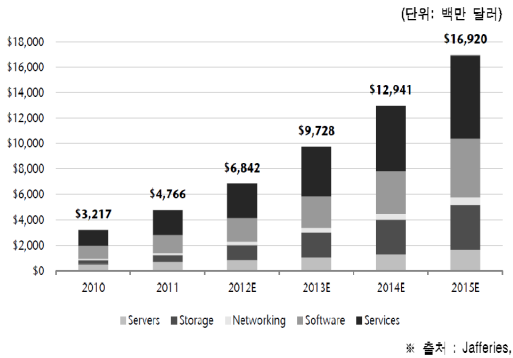


그림 7. IDC의 빅데이터시장 규모전망(2010~2015)
Fig 7. The scale forecast of IDC bigdata market

시장조사업체 IDC는 이 같은 비정형 데이터를 활용하고자하는 업계의 시장수요가 크게 증가하면서, 빅데이터 관련 시장 규모가 2015에는 169억2,000만 달러 수준까지 성장할 것으로 전망하고 있다. 또한 관련 시장은 그림 7 에서 관련시장을 크게 서버, 스토리지, 네트워킹, 소프트웨어, 서비스 부문으로 구분되어 있음을 볼 수 있다. 2011년부터 2015년까지 연평균성장률(CAGR) 34.2%를 기록 하며 2015에는 46억 달러 규모의 시장을 형성할 것으로 내다보고 있다^[3].

그러므로 지식기반경제의 흐름이 제1에 벽에 해당되던 비용의 벽을 넘어, 다음 관문인 축척된 데이터를 비즈니스 인사이트로 도출하여야한다는 것이다.

3. 데이터변화와 힘의 이동

SNS의 일상화가 이뤄지고 라이프로그 데이터등이 결합되어 정보사회의 새로운 패러다임이 형성되어가고 시점에 있다고 하겠다. 특히 능동화, 개인화 등 스마트시대 주요 데이터의 활용이 경쟁력을 좌우하며, 그에 따른 가공과 분석이 새로운 상황을 인식 시키고, 문제해결과 미래에 대한 예견을 투명하게 한다고 하겠다. 그러므로 데이터와 소셜의 관계를 경제적인 자산과 경쟁력의 척도로 제시 하려한다.

표 3. 정보사회의 패러다임 변화

Table 3. The paradigm shift of information society

	PC시대	인터넷시대	모바일시대	스마트시대
패러다임 변화	디지털화, 전산화	온라인화, 정보화	소셜화, 모바일화	지능화, 개인화, 사물정보화
IT 이슈	PC, PC통신, 데이터베이스	초고속 인터넷, www, 웹서버	모바일 인터넷, 스마트폰	빅데이터, 차세대PC, 사물네트워크(M2M)
핵심분야 (서비스)	PC, OS	포털, 검색엔진, Web2.0	스마트폰, 앱서비스, SNS	미래전망 상황인식 개인맞춤형 서비스
대표 기업	MS, IBM 등	구글, 네이버, 유튜브 등	애플, 페이스북, 트위터 등	?
IT비전	1인 1PC	클릭 e-Korea	손안의 PC, 소통	IT everywhere, 신가치창출

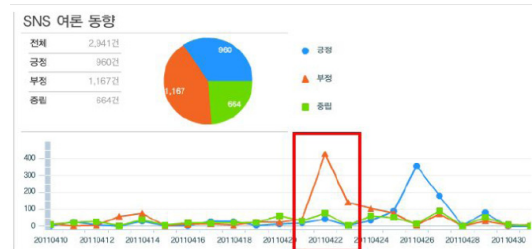
* 출처 한국정보화진흥원(2011.12)

표 4. 빅 데이터의 4가지 구성요소

Table 4. The four kinds of component about big data

구분	주요 내용
규모(Volume)의 증가	• 기술적인 발전과 IT의 일상화가 진행되면서 해마다 디지털 정보량이 기하급수적으로 폭증 => 제타바이트(ZB) 시대로 진입
다양성(Variety) 증가	• 로그기록, 소셜, 위치, 소비, 현실데이터 등 데이터 종류의 증가 • 텍스트 이외의 멀티미디어 등 비정형화된 데이터 유형의 다양화
복잡성(Complexity) 증가	• 구조화되지 않은 데이터, 데이터 저장방식의 차이, 중복성 문제 등 • 데이터 종류의 확대, 외부 데이터의 활용으로 관리대상 증가 • 데이터 관리 및 처리의 복잡성이 심화되고 새로운 기법 요구
속도(Velocity) 증가	• 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

* 출처 한국정보화진흥원(2011.12)



*출처: 강학주(2011), 'Social Big Data & Collective Intelligence'

그림 8. 소셜 데이터분석사례

Fig 8. the analysis practices of social data

그림 8 에서 보는바와 같이 2010년 이후부터 스마트모바일 확산과 함께 사용자들이 자발적으로 참여하고, 정보를 생성하는 소셜데이터 혁명이 발생되고 있다. 즉 정보의 생성과, 규모 그리고 효과 등에서 1990년대에 기업들이 고객의 정보를 축척했던 혁명과 구분하여 새로운 기반의 페이스북, 트위터 등 SNS이용확산과 커뮤니케이션 방식의 변화는 데이터 변혁을 가치화하는 새로운 패러다임으로 재 인식 할 수 있다는 것이다.^[5] 그러므로 기

업은 고객과 시장이라고는 공간을 3차원적인 공간을 새로운 형태로 소통하며, 지식정보와 함께 정서적이고, 감성적인 정보를 가치화 할 수 있고, 더욱이 개인적인 정보가 진실성과 진정성을 감안하는 축적된 정보의 재생산은 향후 IT분야에 새로운 변화의 축을 형성 할 수 있다는 것이다.

IV. 향후전망 및 평가

빅 데이터에 관련한 정부의 새로운 정책들과 관련 대책과제도 발표되고 있다. 또한 2011년 11월 정부는 다양한 데이터를 정책에 활용하는 빅 데이터를 활용한 스마트 정부 구현 방안을 발표했다. 이 방안에 따르면, 정부는 공공과 민간에서 축적한 광범위한 지식 정보를 누구나 활용할 수 있도록 국가지식정보 플랫폼과 민간지식정보 유통을 촉진하는 기능의 사이버 지식정보 거래소를 2013년부터 단계적으로 구축하고, 범국가적 지식정보 활용 촉진에 필요한 법, 제도 개선과 안전한 지식정보 공유를 강화할 예정이다.^[5]

그러므로 빅데이터의 분석기술을 응용할 수 있었던 데이터마이닝기술과 더불어 소셜네트워크기술을 접목하여 개념에서 유형분석을 하고 다양한 시각을 통하여 과거의 데이터를 현재에 대한 문제 해결방안과 각종 사고나 기사를 통하여 무형과 유형 그리고 융합된 형태의 분석기술들을 평가하고 이를 다시 대용량화 되어있는 기업과 개인, 그리고 금융과 개인등의 소통을 통하여 구체화된 분류기술과 지능형서비스까지도 향후제시 할 수 있다고 하겠다.

V. 결론

본 연구는 많은 산업분야에 사용되는 빅데이터, SNS 콘텐츠 및 멀티미디어 등 비정형 데이터 분석 방안을 목적으로 진행 하였다. 그 중에도 빅데이터가 IT업계의 핫이슈로 부상하면서 facebook이나 Tweet와 같은 소셜네트워크서비스(SNS)콘텐츠도 활용도를 높일 필요가 있는 비즈니스 분석의 하나로 부상하고 있다. 신규서비스 개발 시 SNS를 활용하면 기존의 설문조사를 오랜 시간의 소모를 빠른 시간안에 소비자들의 선호도를 파악할 수

있다는 것이다. 최근 사회적으로 활성화되고 있는 동영상이나 사진과 같은 비정형화된 데이터분석도 활용되고 있는 추세이다. 또한 스마트폰을 단순히 핫스팟용으로 이용할 것이 아니라, 각종 관련 평가기술 개발 및 표준화를 위한 기반 자료로 제공될 것이다. 뿐만 아니라 관련기술 외의 빅데이터와 접목된 다양한분석 개발 연구 활성화에 한 역할을 할 수 있을 것으로 보인다. 특히 맥킨지는 비즈니스 지형을 바꿀 10가지 기술 트렌드 중 하나로 빅데이터 선정하였으며, 빅데이터를 수집, 저장하고 이를 토대로 새로운 정보를 찾아내는 것이 경제성장을 위한 중요한 가치창출 효과를 가져 온다고 분석하였다. 끝으로 빅데이터의 향후전망이나 변화의 추이 등을 감안하여 볼 때 빅데이터 환경하에서 필요에 따른 가공과 분석은 소셜의 관계를 더욱더 경제적인 자산과 경쟁력의 척도로 제시 할 수 있다.

참고문헌

- [1] Jyung Hyun Kim, "BigDataplatform-basedsocialnet workdataanalysispractices". 2012.
- [2] KISA Internet Weekly 2012.08.02
- [3] Seung-ku Hwang, "BigDataTechnology Statusand Prospects", 2012. 4. 13.
- [4] Jung Sook Kim, "Big data Utilization and related Technique and Technology Analysis", Vol. 10 No.1
- [5] Tyler Bell(2011), "Big Data: An opportunity in search of a metaphor", radar.oreilly.com/2011/02/big-data-metaphor.html
- [6] Moon Hee Choi, "Research on Utilization of Boryeongmud Marketing through SNS(Social Network Service) Marketing", Journal of the Korea Academia-Industrial cooperation Society, pp.169-172, May 2012.
- [7] H. Yoo, M. Ka, B. Kim, D. Cho, J. Kim, "A Research and Analysis of Use Purpose and Preference of Smart Phone SNS", Journal of the Korea Academia-Industrial cooperation Society, pp.235-237, Dec 2011.

※ 본 논문은 2012년도 한세대학교 학술연구비 지원에 의해서 연구되었음

저자 소개

신 승 중(정회원)



- 1988년 : 세종대학교 경영학과 졸업 (경영학석사)
- 1994년 : 건국대학교 전자계산학과 졸업(공학석사)
- 2000년 : 국민대학교 정보관리학과 졸업(공학박사)
- 2008년 ~ 현재 : 한세대학교 IT학부.

<주관심분야 : 정보관리, 정보전, 정보보호>