



특집 10

도큐먼트형 빅데이터의 분석을 이용한 비즈니스 인텔리전스



권영현 · 강용성 · 안영민 (㈜와이즈넷)

목 차 »	1. 서 론
	2. 관련 연구
	3. 도큐먼트형 빅데이터를 이용한 BI
	4. 결 론

1. 서 론

앨빈 토플러(Alvin Toffler)가 1980년, 제 3의 물결^[1]에 대한 화두를 던진 이후, 폭발적으로 증가하는 데이터의 홍수 속에서 30 여 년의 시간이 흘렀다. 시대를 막론하고 데이터의 중요성은 항상 존재해왔으나, 현재의 인터넷과 모바일 환경을 고려해보면, 데이터의 생성 양상이 다양해지고 그 분석은 더욱 중요하게 되었다. 예전부터 데이터의 크기가 기하 급수적으로 증가하고 있다는 것을 주시하던 선구자들은 큰(Large Scale), 매우 큰(Very-Large Scale) 데이터 등의 용어를 사용하여 일반적인 방법으로 처리할 수 없는 데이터를 표현해 왔다. 하지만 최근에 급부상 하고 있는 빅데이터의 개념에 대해 Terence와 Mary^[2]는 단순히 가공할 규모뿐만 아니라 수많은 다양성을 내포하고 있는 사용자 제작 콘텐츠(User Generated Contents) 라고 표현했다.

이러한 빅데이터가 공간적, 문화적 제약을 초월해서 매우 큰 영향력을 행사함에 따라 빅데이

터라는 용어는 단순히 데이터 크기를 표현하는 용어라기 보다는 하나의 거대한 기술, 문화, 사회 전반을 아우르는 새로운 문명의 태동으로 파악되고 있다. 때문에 학계는 물론 기업, 기관에 이르기까지 빅데이터 내부의 숨겨진 개념, 가치, 의미 등을 규명하려는 노력과 요구가 생겨나고 있다.

이러한 요구들이 생겨나는 이유는 빅데이터의 특징 때문이다. 기존의 기술들은 학계에서 발표된 후 실용화되어 기업 및 기관에서 이용되는 것이 일반적인 흐름인 반면, 빅데이터에 대해서는 기업이 더 민첩하고 적극적으로 연구에 참여하고 있다. 이는 현실적 데이터 폭증에 당면한 기업들의 데이터 관리 요구에 따른 반작용이라고 볼 수도 있겠지만, 한편으로는 빅데이터 분석에 따르는 실리(實利)를 발견한 것이라고 볼 수 있다. 상투적 표현으로, 기업은 이윤을 추구하고 투자 가치가 있다고 판단되는 분야에는 적극적인 행보를 보이는 것은 당연하기 때문이다.

이러한 요구에 따라 연구되고 있는 빅데이터를 이용한 사업적인 분야는 크게 두 가지로 구분해

볼 수 있다.

첫째로 전사적 자원관리(ERP), 기업지식포탈(EKP) 등의 분석을 통해 기업의 의사결정을 돕는 DSS^[3] (Decision Support System)와 같이 전통적인 형태가 있을 수 있으며, 둘째로는 엄청난 양산 속도를 가진 SNS 와 사용자 생성 콘텐츠(UGC)의 정성적인 분석을 통해 소비자의 트렌드와 피드백을 분석/연구하여 전략적인 의사결정을 하기 위한 비즈니스 인텔리전스(New Business Intelligence)가 있을 수 있다.

위의 두 분야는, 비선형 양산 속도를 가진다는 점과 비정형 데이터에 초점이 맞추어져 있다는 점에서, 기존의 데이터 관리 소프트웨어의 수집, 저장, 관리, 분석의 역량을 초월하는 크기를 가진 데이터^[4]라고 표현되는 빅데이터의 정의에 부합한다.

또한 다른 시각으로 보면 데이터의 성격에 따라 구분 할 수가 있는데, 그 성격에 따라 분석 방법이 달라질 수 있다. 크게 다음과 같은 데이터의 성격과 그에 따른 방법론들이 있을 수 있다.

먼저 데이터 자체가 의미가 있어 인간의 인지 능력으로 시간만 있으면 이해가 가능한 도큐먼트 데이터(Document Data)가 있을 수 있으며, 이것

은 그 수량과 데이터 저장형식에 초점이 맞추어진 분야가 있으며, 다른 한편으로는 그 데이터가 가지는 의미와 내용에 초점이 맞추어진 분야가 있다. 이 예에는 SNS, 블로그, 이메일 등의 콘텐츠에 해당하는 내용이 있는데, 이런 형식의 데이터들은 개별적인 데이터가 가지는 가치뿐만 아니라 대량으로 형성되었을 때, 집단지성으로 대변되는 그 가치와 의미가 배가 된다고 볼 수 있다.

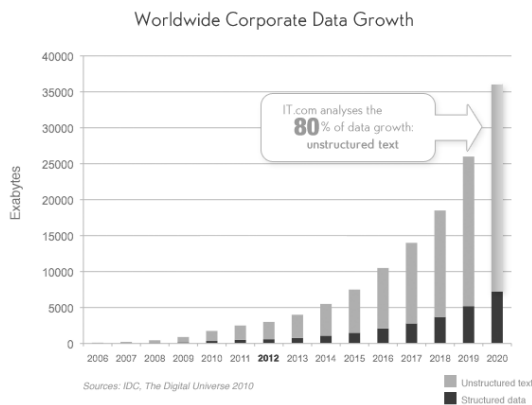
둘째로는, 비도큐먼트 데이터(Non-Document)가 있다. 단일 단위로는 그 의미를 가지기 어렵고, 군집이 되었을 때, 특정 패턴을 통해 숨은 의미가 발견될 가능성이 있는 데이터이다. 이 예에는 쇼핑 패턴, 생산라인에서의 센서 정보(온도, 습도, 압력 등), 위치정보 등을 이용한 사례가 있다.

이 데이터는 대량의 군집 데이터에 존재하는 숨은 의미를 찾는 것에 분석 의의가 있다는 것이 공통점지만, 본 논문에서는, 주로 도큐먼트 데이터의 분석 기법을 소개하는 동시에 활용방안을 모색하고, 그 구체적인 사례를 들어 빅데이터 시대의 집단지성 분석을 통한 비즈니스 인텔리전스의 모델을 제시하고자 한다.

2. 관련 연구

빅데이터의 활용을 위해 주안점을 두어야 하는 부분에 대해 Doug^[6]는 IT리더가 집중해야 하는 빅데이터의 특징 3가지를 3V로 표현했다. 첫째로 전통적인 데이터 타입과 더불어 새로운 타입의 데이터를 포괄하며 크기(Volume)가 방대해진 것이고, 필연적으로 그에 비례하는 다양성(Variety)을 가지고 있으며, 데이터 증가 속도(Velocity)가 빠르다는 점이 그것이다.

위 특징과 더불어 빅데이터의 처리의 활용을 위해서는 필수 불가결하게 강력한 컴퓨팅 파워가 요구되는데, 일반적인 단일 기기로서 대량의 데



(그림 1) 비정형 데이터의 증가 추이^[5]

이터를 처리할 수 있는 한계를 넘어선 현 상황에서 제시된 대안은 분산 처리 환경의 요구였다. 이러한 시대적 요구에 의해, 구글 GFS, 하둡 HDFS 등의 분산 파일 시스템 환경이 주목을 받기 시작했고 단순한 분산 프로그래밍 프레임워크를 넘어 기존 RDBMS의 변형 형태인 NoSQL이 이슈화되기 시작했다. 최근에 급속도로 주목 받고 있는 NoSQL은 Brewer에 의해 제시된 CAP 이론^[7] (CAP Theorem) 중 일관성(Consistency), 혹은 사용성(Availability)보다 분산 수용성(Partition Tolerance)에 초점을 두어 데이터의 수평적 확장에 용이성에 의의가 있는 데이터 저장/관리 형태이다. Gilbert and Lynch^[8]에 의해 증명된 CAP 이론에 따르면 NoSQL은 CAP 3개의 개념을 동시에 충족시킬 수 없다고 했으며, 이에 따라 모든 기존의 RDBMS의 역할을 대신할 수 없는 것은 자명한 사실이다. 비단 NoSQL 뿐만 아니라 어떠한 데이터 분산 저장/관리 형태의 구조라 할지라도 CAP 이론의 제약을 받으므로, IT 리더 뿐만 아니라, 시스템 관리, 구축, 개발에 연루된 모든 사람들은 데이터의 성격과 목적을 분명히 알아야만 빅데이터에 대해 제대로 인식하고 있다고 할 수 있다.

데이터 저장/관리 형태뿐만 아니라 분석은 기업의 의사결정을 위한 가장 중요한 요소라고 할 수 있다. 분석 대상을 정해야 목적과 방향을 분명

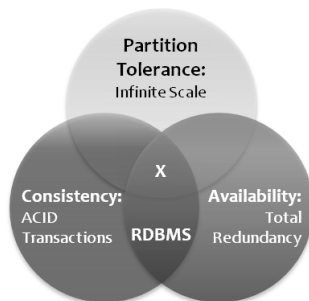
히 할 수 있으므로 분석 대상을 크게 비도큐먼트 데이터와 도큐먼트 데이터로 나누었다. 서론에 언급한대로 우리의 주요 분석 대상은 단일 데이터로도 의미를 담고 있고, 군집의 형태를 이루었을 때에도 그 추출해야 할 의미가 산재해 있는 도큐먼트 데이터(Document Data) 형태의 분석에 대해 다룰 것이다.

필연적으로 우리의 연구는 텍스트 마이닝(Text Mining)에 밀접한 관련이 있으며 이는 주로 주제 추출(Topic Extraction)과 자연언어처리(NLP)에 근간을 두고 있다. (주)와이즈넷은 이 도큐먼트 데이터의 분석을 위해 Blei et.al의 LDA(Latent Dirichlet Allocation)^[9]를 이용한 토픽 모델을 차용했고, 자연언어처리 부분에 대해서는 김영택^[10], 강승식^[11]의 저서와 안영민^[12] 등을 참조하여 도큐먼트 형 빅데이터 분석에 사용하였다.

3. 도큐먼트형 빅데이터를 이용한 BI

사람이 생성하고 이해할 수 있는 문서형 데이터 집합을 말하며, SNS, 블로그, 이메일 등의 콘텐츠가 이것에 속한다. 블로그와 SNS를 통해 고객 소통하려는 기업의 노력은 이미 오래 전부터 있어왔으나 최근에 들어서야 사회, 집단적으로 발생하는 트렌드나 Buzz 등의 집단 의사에 대해 위 콘텐츠를 이용하는 방법의 연구와 제품화가 이루어졌다. 또한, 비단 기업뿐만 아니라 정치, 경제, 사회, 문화에서도 SNS에서 발생하는 집단 의사 형성에 촉각을 곤두세우고 있으며, 기존 매체 대비 능동적이고 현실적인 의사를 수렴할 수 있으므로 가장 영향력 있는 매체로 자리 잡았다.

본 장에서는 이러한 매체로부터 발생된 데이터를 다양한 텍스트 마이닝 기법을 통해 분석하는 방법을 개념적으로 설명하고, 분석 결과를 토대



(그림 2) CAP 이론의 3 요소와 주요 요소

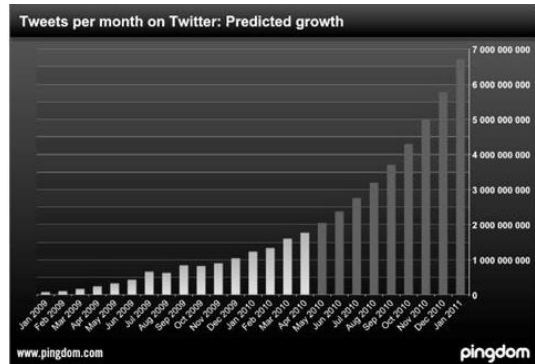
로 기업의 의사결정에 어떻게 도움이 될 수 있는지를 집중적으로 다룬다.

3.1 자연언어 처리를 통한 집단 지성/ 감성의 발견

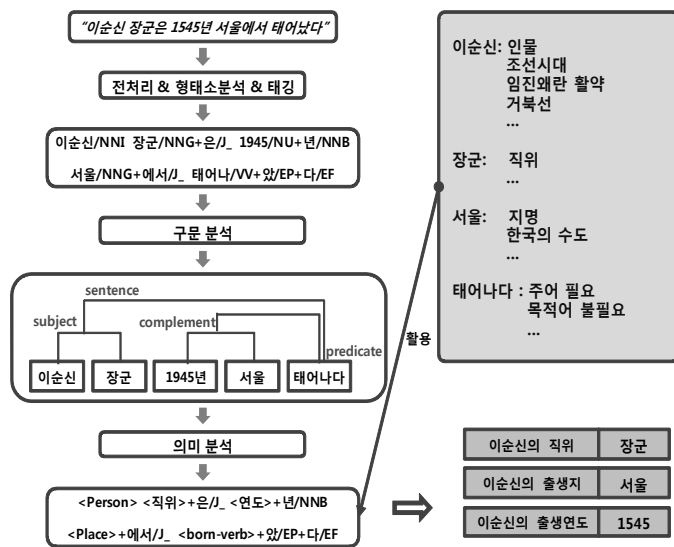
자체 조사에 의하면, 인터넷상에 존재하는 SNS 기반 문서는 정서, 느낌(Sentiment)이 포함된 복수 단어 형태의 의미패턴이 평균 3.2개가 존재한다. 또 2010년 Pingdom의 조사자료^[13]에 의하면 대표적인 SNS 서비스인 트위터(Twitter)의 사용자는 1.45억 명이고, 9000만 트윗(Tweet)이 단 하루에 발생하며, 이 값은 매일 30만 트윗씩 증가하고 있다. 바꾸어 말해, 2010년에 트위터를 통해 분석할 수 있었던 의미 패턴은, 단순히 선형 증가 함수의 누적으로 계산해도 약 1500억 개라고 할 수 있다.

이런 거대한 양의 데이터로부터 생각을 추출할 수 있는 텍스트 마이닝 방법에는 여러 가지 접근법이 있겠지만, 우리가 제안하는 첫번째 방법은 자연언어처리(NLP)를 이용한 의미 분석이다.

의미 분석을 위한 전통적 자연언어처리의 단계



(그림 3) 월별 트위터(Twitter) 트윗(Tweet) 증가 추이는 3가지로 요약된다. 첫째로 원문 문서를 그대로 사용했을 때 발생할 수 있는 중의성을 제거하고, 의미를 가지는 최소단위로 구성 성분을 분해하는 형태소 분석과 태깅 과정(사전을 참조하여 문장을 품사 등의 카테고리로 분해하여 원문에 기술하는 과정), 그리고 형태소의 관계를 규명하여 문장의 구조를 만드는 구문 분석 과정(Syntactic Analysis), 마지막으로, 구조화된 문서를 통해 구조화된 구문들의 의미를 추출하는 의미 분석(Semantic Analysis) 과정이 그것이다. (그림 4)는



(그림 4) 전통적 자연언어처리를 이용한 의미분석기법

위 3가지 과정을 보여준다.

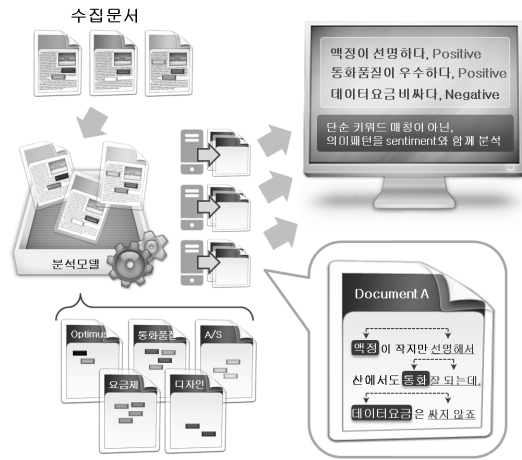
예에서 이순신의 직위, 출생지, 출생연도는 "이순신 장군은 1545년 서울에서 태어났다"라는 하나의 문장이 포함하고 있는 의미를 파악하는데 있어서, 형태소 분석, 태깅, 구문분석, 의미분석에 이르는 과정을 통해 이루어졌다. 그러나 위의 예에서 설명된 전통적 자연언어처리는 실질적 사용에 제약이 따른다. 제약이 따르는 가장 큰 이유는 완전한 구문 분석 및 의미 분석이 아직까지 가능하지 않는 이유로 품질에 확신을 가질 수 없기 때문이다. 그래서 우리는 의미 분석을 적용하기 위한 대상을 명확히하여, LSP(Lexico Symantic Pattern)을 정의한 다음 이를 이용하여 매칭하는 방법론을 제안한다.

LSP^[14]를 사용한 의미 분석 기법은 텍스트의 토큰 혹은 문장의 성분 구조를 패턴화 하여 규칙(Rule)으로 만들어 묶어 하나의 분석모델로 구축한다. 이 구축된 모델(지식)을 통해 대상 전체 문서 집단에 대한 의미를 추출하는 기법을 토큰과 문장 성분 구조에 각각 적용하면, (그림 4)에 설명된 구문 분석과 의미분석을 수행하는데 적용될 수 있다.

이 방법론을 사용하여 기업의 의사결정에 도움을 줄 수 있는 대표적인 적용 분야는 오피니언마이닝(Opinion Mining)을 들 수 있는데, (그림 5)는 이 예를 보여준다.

예제에서는 SNS등을 통해 수집된 문서를 휴대폰에 대한 호불호(好不好)를 판단하기 위해 통화품질, 디자인, 요금제, 제품명 등을 분석모델에 구축하고, 모델 자체를 여러 머신으로 나누어 처리한 후, 결과를 모아 의사 결정권자에게 보이는 모습을 도식했다.

기존의 오피니언마이닝 대안기법들이 전체 소비자의 일부분의 대표를 표본으로 추출하여 설문, 분석하는 정도에 그쳤다면, 문서형태 빅데



(그림 5) NLP를 이용한 오피니언마이닝의 예

이터를 자연언어처리 후 분석하는 기법을 사용하면 표본집단의 크기가 모집단에 상대적으로 근접하게 된 것에 그 의미가 있다고 하겠다. 이를 통해 올바른 의사결정에 필요한 기초 자료의 정확도가 향상될 것이며, 양산속도가 빠른 빅데이터의 특징과 빅데이터 분석으로 발빠르게 시장의 변화에 대처할 수 있을 것이다. 이로써, 어제 출시한 제품을 구매한 10만 명이 생산한 지배적 피드백을 오늘 받을 수 있는 것이 요원한 것만은 아니게 되었다.

3.2 주제 추출 기반 분석 기법

본 논문에서 제시하는 문서형태 빅데이터를 이용한 비즈니스 인텔리전스의 두번째 사례는 주제 추출(Topic Extraction)로서, Thomas Hoffan이 1999년 PLSI^[15](Probabilistic Latent Semantic Indexing)을 이용한 토픽 모델을 선보인 이후, 현재에 가장 많이 활용되는 LDA에 이르기까지 많은 연구가 이루어진 분야이다. 우리는 주제 추출 기법 자체에 초점을 맞추기 보다, 그것이 무엇을 하기 위한 도구이며, 무엇을 할 수 있을지에 초점을 맞추었다.

주제 모델(Topic Model)이란, 문서 집합(Corpora) 내의 숨겨져 있는 관계를 수학적으로 규명하고, 문서들에 그 토픽들을 표기해서 그 데이터를 체계화, 요약하고 검색하는 제공하기 위한 모델이다. 여기서 주제 추출을 한마디로 정의한다면, 이러한 주제 모델을 사용하여 분류(Unlabeled)되지 않은 거대한 문서 집합에서 비슷한 주제를 가진 문서들을 군집화(Clustering) 하는 것을 말한다고 할 수 있다.

현재 알려진바로는 가장 성능이 좋은 LDA는 생성 확률 모델(Generative Probabilistic Model)을 사용하여 주제의 개수와 분포 등의 파라미터를 알고 있다고 가정한 상태에서의 문서의 특성을 모델링 한 다음에, 역으로 실존하는 문서의 모델링을 정확하게 했다고 가정했을 때의 파라미터를 추론, 활용하여 주제를 군집화하는 방식이다. (그림 6)은 이렇게 주제 모델을 통해 추출된 결과를 보여준다.

이러한 방법론을 통해 추출된 개개의 주제어들은 해당 분야 저자들의 선호도가 높은 단어들로 이루어진다고 할 수 있고, 주제는 주제어를 대표한다고 할 수 있다. 이를 통해 집단 감성의 분류가 가능하며, 나아가 현재의 트렌드를 읽을 수도 있다. 또 단순히 문서 집단에서 주제별로 문서를

"Arts"	"Budgets"	"Children"	"Education"
new	million	children	school
film	tax	women	students
show	program	people	schools
music	budget	child	education
movie	billion	years	teachers
play	federal	families	high
musical	year	work	public

The William Randolph Hearst Foundation will give \$1.25 million to Lincoln Center, Metropolitan Opera Co., New York Philharmonic and Juilliard School. "Our board felt that we had a real opportunity to make a mark on the future of the performing arts with these grants an act every bit as important as our traditional areas of support in health, medical research, education and the social services," Hearst Foundation President Randolph A. Hearst said Monday in announcing the grants.

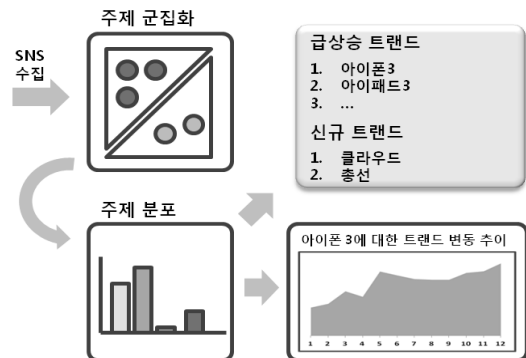
(그림 6) 주제 추출의 예

군집할 수 있는 것뿐만 아니라, 관심 주제, 전문 분야로 그 문서의 저자들을 군집화 하여 추천 시스템(Recommendation System)이나 연관 인물 검색 등에 활용될 수 있다.

주제 추출 기법을 사용한 기업의 의사결정을 돕는 시스템의 사례로는 트렌드 분석과 연관 인물 검색을 대표적인 두 가지 사례로 들 수 있다.

첫째로 주제 추출을 사용하는 트렌드 분석 기법은 군집화 이후에 파생되는 주제 분포(Topic Distribution)를 주기적으로, 일정 시간 간격으로, 그 변량을 분석하여 해당 주제에 대한 관심 변화를 보여주는 트렌드 변동 분석이 있으며, 현 시점에서의 Top K를 순위화 하여 트렌드 순위를 매기는 트렌드 순위 분석 등이 있다.

(그림 7)은 설명된 트렌드 분석 기법에 예시를 도식한 것이다. 이 예에서는 SNS로부터 수집된 데이터를 사용하여, 주제 모델로부터 군집화를 수행하고, 주제 분포를 구하는 과정을 일정 시간 간격으로 반복했다. 이 반복된 결과를 토대로 이전 결과와 비교하여 특정 제품의 트렌드 변동 사항을 보여주고 있으며, 특정 시점의 트렌드 랭킹을 분포의 Top K를 추출하여 순위 매김 하거나, 새로 진입 되었으면서도 분포의 큰 부분을 차지하는 신 트렌드에 대한 순위 매김을 하고 있다.



(그림 7) SNS를 이용한 트렌드 분석

또 다른 활용 사례로, 주제 추출을 통하여, 주제를 군집화하고, 같은 주제에 대해 많은 콘텐츠를 생성한 인물을 연관 인물로 인식하는 방법과, 주제에 속하는 주제어 사용의 빈도에 따라 저자가 속하는 관심분야를 추정하여 같은 관심분야를 가진 사람을 추천하는 방법이 있을 수 있다.

이 연관 인물 분석 기법을 사용하여 활용될 수 있는 기업 내부의 적용 분야를 알아보면, 인사 분야에 종사하는 사람은 전문, 관심 분야가 같은 사람을 가급적 같은 부서로 배치하는 전략적인 의사결정의 근간 데이터가 될 수 있으며, 또 다른 활용으로는, 기업에서 양산된 문서 집단이 기업의 방향성을 표현한다고 가정할 수만 있다면, 높은 비율의 주제에 대해 긍정도/부정도를 판단하여 해당 주제에 속하는 인원을 제재하거나 칭찬할 수 있을 것이다. 개발부서에서는 전문분야를 나누어 효율적으로 업무를 분담할 수도 있고, 경영자가 이를 사용한다면, 기업 내 커뮤니케이션의 흐름을 진단하여 흐르지 않고 병목 되는 구간을 살피고 해소 방안을 마련할 수 있을 것이다.

이와 같이, 기업에서 빅데이터의 주제 추출을 활용할 분야는 기업 내부의 데이터는 물론 외부의 데이터의 분석에 이르기까지 다양하다. 더군다나, 본 장에서 제시된 활용분야는 극히 일부분에 불과하며, 지금의 빅데이터 개념의 태동의 시기를 지나게 되면 더욱 더 많은 활용 분야와 기법들이 연구되어 생성될 것이다. 앞으로 빅데이터의 분석을 통하여, 기업은 더욱 지능적 경영을 하게 될 것이며, 인적, 물적 자원에 대한 지출을 줄이면서도 정확한 트렌드를 짚어낼 것이다. 빅데이터를 통한 기업의 예측 및 의사결정은 또 한번의 지식 경영 고도화 시대를 예고한다.

4. 결론

본 논문에서는 새로운 트렌드로 자리매김하고 있는 빅데이터 시대의 흐름에 맞추어 도큐먼트형 빅데이터 처리, 분석의 전망과 그 가능성에 대해 살펴보았다. 우리는 빅데이터의 정의부터, 형식과 내용적, 의미적 기준에 따른 분류방법을 소개했고 그들이 가진 특징들을 연구했으며, 특히 도큐먼트형 빅데이터 분석에 있어 자연언어처리 및 주제 추출 방법론을 적용할 수 있는 분야를 기술했다. 뿐만 아니라, 소개된 분석 방법론을 통해 얻어질 수 있는 기회 내지는 이득을 기업의 시각으로 바라봄으로써, 보유하고 있거나, 취득할 수 있는 빅데이터를 활용한 기업의 실리를 구체화하였다. 이를 통해 데이터를 직접 활용할 기업뿐만 아니라 기관 및 학계에도 새로이 정의된 패러다임에 적용할 수 있는 현실적 대안(對案)에 대한 선례(先例)를 보였다고 할 수 있다.

참고 문헌

- [1] Alvin Toffler, "The Third Wave", Bantam Books, 1980.
- [2] Terrence Craig, Mary E. Ludloff, "Privacy and Big Data", O'Reilly Media, 2011.
- [3] Peter G. W. Kenn, "Decision Support Systems: An Organizational Perspective", 1978.
- [4] Wikipedia, "Big data", 2011, http://en.wikipedia.org/wiki/Big_data
- [5] it.com, "World Corporate Data Growth", 2010, <http://www.it.com/solutions/>
- [6] Doug Laney, "3-D Data Management: Controlling Data Volume, Velocity and Variety", META Group Inc., 2001, <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>

- [7] Eric A. Brewer, "Towards robust distributed systems", Proceedings of the nineteenth annual ACM symposium on Principles of distributed computing, p.4, 2000.
- [8] Nancy Lynch, Seth Gilbert, "Brewer's conjecture and the feasibility of consistent, available, partition-tolerant web services", ACM SIGACT News, Vol.33, No.2, pp.51-59, 2002.
- [9] David M. Blei, Andrew Y. Ng, Michael I. Jordan, "Latent Dirichlet Allocation", The Journal of Machine Learning Research, Vol.3, pp.993-1022, 2003.
- [10] Yung Taek Kim, "Natural Language Processing", Life & Power Press, 2003.
- [11] Sung-sik, Kang, "Han'gugo hyongt'aeso punsok kwa chongbo komsaek", Hongneung Kwahak Ch'ulp'ansa, 2003.
- [12] Young-Min Ahn, Su-Hyun Oh, Yu-Hwan Kang, Young-Hoon Seo, "Answer Extraction of Concept based Question - Answering System", Vol.2, No.1, pp.448-451, 2005.
- [13] Pingdom, "How much we will be tweeting by 2011", 2010, <http://royal.pingdom.com/2010/05/07/how-much-we-will-be-tweeting-by-january-2011-chart/>
- [14] Paul S. Jacobs, George R. Krupka, Lisa F. Rau, "Lexico-Semantic Pattern Matching As Companion To Parsing In Text Understanding", Proceedings of the workshop on Speech and Natural Language, pp.337-341, 1991.
- [15] Thomas Hofmann, "Probabilistic Latent Semantic Indexing", Proceedings of the 22nd International Conference on Research and Development in Information Retrieval(SIGIR), 1999.

저 자 약 령



권 영 현

이메일 : lfrost@wisenet.co.kr

- 2008년 건국대학교 인터넷학과 (학사)
- 2010년 건국대학교 신기술융합학부 (석사)
- 2010년~현재 ㈜와이즈넷 연구원
- 관심분야: 검색엔진, 클라우드컴퓨팅



강 용 성

이메일 : scott@wisenet.co.kr

- 1997년 건국대학교 철학과 (학사)
- 2001년 LAS 21 독일법인 Deuche LAS GmbH 법인 책임
- 2001년~현재 ㈜와이즈넷 혁신기획실장 / 상무
- 관심분야: Knowledge Base, Business Intelligence



안 영 민

.....
이메일 : maniac@wisnut.co.kr

- 2000년 충북대학교 컴퓨터공학과 (학사)
- 2002년 충북대학교 컴퓨터공학과 (석사)
- 2008년 충북대학교 컴퓨터공학과 (박사)
- 2009년~현재 ㈜와이즈넷 연구소장
- 관심분야: NLP & Data mining