

마케팅논집
제21집 제2호 통권 58호

ISSN : 1229-3180(Print)

빅데이터 분석의 기술마케팅 활용에 관한 연구 : 잠재 수요기업 발굴을 중심으로

전채남, 서일원

To cite this article : 전채남, 서일원 (2013) 빅데이터 분석의 기술마케팅 활용에 관한 연구 : 잠재 수요기업 발굴을 중심으로, 마케팅논집, 21:2, 181-203

- ② earticle에서 제공하는 콘텐츠를 무단 복제, 전송, 배포, 기타 저작권법에 위반되는 방법으로 이용할 경우, 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

www.earticle.net

빅데이터 분석의 기술마케팅 활용에 관한 연구 : 잠재 수요기업 발굴을 중심으로*

전 채 남 · 서 일 원

요 약

최근 빅데이터에 대한 연구가 늘어나고 있지만 대부분 빅데이터의 개념, 동향, 기술 현황, 활용 가능성 등의 연구에 국한되어 있다. 본 연구는 빅데이터 연구를 확대하고 실용성을 높이기 위해 마케팅 분야의 빅데이터 활용을 연구 주제로 선정하였다. 빅데이터의 활용 분야로 기술마케팅을 선택한 것은 기술시장에서 마케팅의 필요성이 높아지고 특히 기술시장의 특성 때문에 잠재 수요자 발굴이 중요하기 때문이다. 이를 위해 신기술 빅데이터의 분석을 통해 확인된 주요기술은 무엇인가?, 신기술 빅데이터의 분석을 통해 발굴된 잠재 수요기업은 어디인가?, 잠재 수요기업과 연관되어 있는 핵심어들은 무엇인가? 등을 연구문제로 설정하였다.

한국표준과학연구원(KRISS)의 2가지 신기술을 선정하고 3명의 코더들이 협의를 통해 4개의 분석단어를 정하였다. 빅데이터는 인터넷에서 수집하고 정제 한 후에 텍스트마이닝과 시맨틱네트워크분석을 실시하였다. 기업의 스키마와 홈페이지 통해 재검증하는 과정을 거쳤다. 빅데이터 분석을 실시한 결과, 첫 번째 연구문제는 연결정도, 연결정도중심성, 빈도 등을 통해 2개의 기술마다 상위 20개의 주요기술을 확인하였다. 두 번째 연구문제는 공동출현 연결망, 연결 강도, 코사인 유사계수 등을 통해 2개 기술의 잠재 수요기업을 발굴하였다. '배열검출기형 분광복사계 성능평가 기술'은 17개의 기업을 발굴하였고, '초음파를 이용한 진공압력 측정센서 및 모듈'은 10개의 기업을 발굴하였다. 세 번째 연구문제는 잠재 수요기업의 에고네트워크분석을 통해 기업명과 의미 있게 연관되어 있는 단어들을 확인하였다.

본 연구는 기존의 개념적 빅데이터 연구에서 빅데이터 분석을 통해 기술마케팅의 잠재 수요기업을 발굴하는 실용적인 빅데이터 연구로 나아간 점이 의의가 있다. 빅데이터의 구체적인 활용 분야로 기술마케팅을 선택하여 실증함으로써 연구의 범위를 확대하고 향후 빅데이터 분석 연구의 발전에 기여하고 있다.

주제어 : 빅데이터, 빅데이터 분석, 시맨틱네트워크분석, 기술마케팅, 수요기업 발굴

* 이 논문은 아시아트리플헬릭스학회의 춘계세미나 발표자료인 『빅데이터 분석의 활용 방안에 관한 연구:기술마케팅의 잠재 수요자 발굴을 중심으로』를 참석자들의 조언과 추가 연구를 통하여 수정하였습니다.

전 채 남 더아이엠씨(THE IMC), 대표이사, cnjun@theimc.co.kr (주저자, 교신저자)

서 일 원 한국표준과학연구원(KRISS) 기술사업화센터(TLO), veny@kriss.re.kr (공동저자)

논문접수(2013.06.07), 논문심사(2013.06.25), 게재확정(2013.07.01)

I. 연구배경 및 목적

인터넷의 생활화에 이은 모바일 디바이스의 대중화와 소셜네트워크서비스(SNS)의 확대로 사람들이 이용하는 미디어와 커뮤니케이션의 영역이 확장되고 있다. 그 결과 엄청난 양의 정보 데이터들이 시시각각 생산되고 있다. 시장조사 전문기관인 IDC의 ‘전 세계 빅데이터 기술 및 서비스 전망보고서’(2012, 3)에 의하면 2011년 한 해에만 1.9ZB(제타바이트, 1조 9억천억 기가바이트)이상의 데이터가 생산되거나 복제되었고 향후 5년 내에 거의 9배까지 증가할 것으로 예측하고 있다. 바야흐로 데이터 산맥의 시대, 빅데이터의 시대가 열리고 있다.

빅데이터는 다양한 디바이스와 서비스의 이용자들의 의해 실시간으로 생산된 데이터로 데이터의 양이 많고(Volume) 종류가 다양하며(Variety) 집적속도가 매우 빠르(Velocity) 특성을 가지고 있다. 현재 빅데이터는 특정 분야에 국한되지 않는 유망 산업 분야로 빅데이터 처리 및 분석 능력은 기업의 미래 경쟁력으로 인식되고 있다.

세계경제포럼은 2012년 가장 주목할 기술로 빅데이터를 지목하면서 데이터 과잉 문제를 해결하고 데이터를 활용하는 것을 현안으로 선정하였다. 빅데이터는 대부분 텍스트, 이미지, 동영상 콘텐츠 등의 형태로 기존 방법으로는 데이터를 수집하고 분석하기 힘든 비정형 데이터(Unstructured data)가 많기 때문에 활용이 매우 어렵다. 이와 같은 이유로 최근 빅데이터에 대한 연구들은 대부분 빅데이터 개념, 동향, 기술 현황, 활용 가능성 등의 연구에 국한되어 있다(강안모 외, 2012; 김정숙, 2012; 김한나, 2012; 이만재, 2012; 안창원, 황승구).

빅데이터 연구의 폭을 개념과 현황 소개 단계에서 벗어나 실질적인 응용단계로 제고하기 위해서 본 연구에서는 기술마케팅 분야를 선정하였다. 기술마케팅은 R&D에 의해 창출된 특허, 기술 등의 지식재산권을 사업화하고자 하는 기업에 이전하기 위한 기획과 실행 과정이다. R&D 4.0 시대를 맞아

R&D와 사업화가 융합되면서 수요자의 다양한 욕구를 파악하여 시장의 공급자와 수요자의 미연결을 줄이기 위해 기술마케팅이 더욱 강조되고 있다. 그러나 기술마케팅은 유형의 제품이나 서비스를 일반 소비자와 교환하는 일반 마케팅과는 상이한 점들이 많다. 제품으로서 지식재산권은 무형이면서 미래형이기에 수요자가 제한되어 있다. 마케팅 대상으로서의 지식재산권의 특징은 수요자에 따라 대상의 활용범위가 달라지기 때문에 광범위한 범위에서의 수요 파악과 함께 활용가능성과 의지가 있는 기업을 동시에 탐색해야한다는 점이다. 이와 같은 차이점 때문에 전략적인 기술마케팅은 기술을 필요로 하는 목표 집단을 과학적으로 찾아내는 잠재 수요자 발굴이 매우 중요하다.

현재까지는 기술마케팅이 과학적으로 정립되지 못하고 개별 기술개발자들의 네트워크, 오프라인 활동 등 수공업적 기술이전 활동이 주를 이루고 있는 실정이다. 빅데이터의 활용 분야로 기술마케팅을 선택한 것은 기술마케팅의 핵심인 잠재 수요자 발굴에 효과적인 수단으로 여겨지기 때문이다.

효과적으로 기술마케팅의 잠재 수요자를 발굴하기 위해 새로운 데이터와 분석 방법이 이용되는 빅데이터를 활용한다는 것은 방대한 양의 데이터를 수집하여 처리하고 분석하는 기술을 필요로 한다. 이를 위해 데이터의 창고인 인터넷에서 기술관련 핵심어를 통해 데이터를 수집하고 처리한 후 텍스트마이닝과 시맨틱네트워크분석을 통해 기술이전 잠재 수요기업을 발굴하고자 한다. 이를 위해 빅데이터의 텍스트마이닝과 시맨틱네트워크분석이 잠재 수요기업 발굴을 위한 과학적 방법인지를 문헌연구를 통해 먼저 살펴보고 기술마케팅의 잠재 수요기업 발굴 모형을 새롭게 모형을 제시한다. 이를 토대로 분석 대상 및 범위를 정하여 데이터를 수집하고 빅데이터 분석 절차와 방법에 따라 분석을 실시한다. 이와 같은 실증 과정을 통해 빅데이터 분석의 활용 가능성을 알아보고자 한다.

본 연구는 빅데이터 연구를 개념적이고 현황 소

개 수준에서 실용적 수준까지 확대해 보는 점과 기술마케팅의 잠재 수요기업 발굴의 새로운 방법으로 빅데이터 분석을 활용해 본다는 데 의의가 있다. 빅데이터의 구체적인 활용 분야로 기술마케팅을 살펴보고 빅데이터 연구를 확대함으로써 향후 빅데이터 연구의 발전에 기여할 뿐만 아니라 표적집단 설정의 새로운 방안을 제시하여 기술마케팅의 과학적인 전략 수립에 기여할 수 있다.

II. 이론적 배경

1. 빅데이터(Big data)의 개념 및 활용

1.1. 빅데이터의 개념

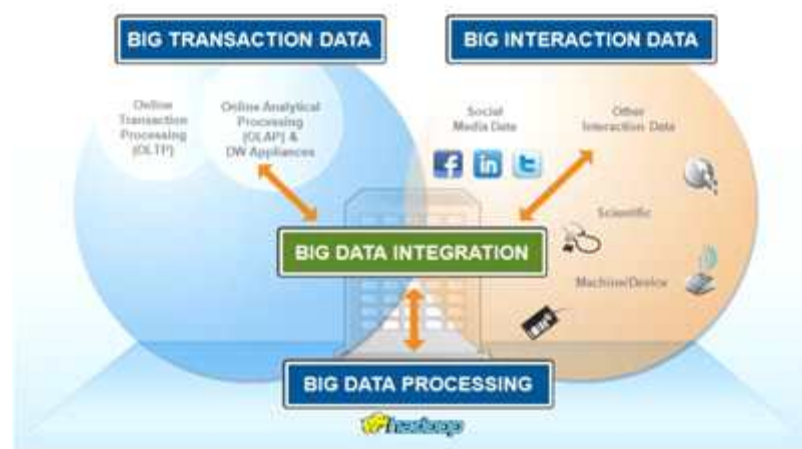
빅데이터(big data)에 대한 정의는 데이터(data)에 빅(big)을 붙인 단어 형태 때문에 주로 데이터의 크기를 주목하지만 다른 특성들도 포함하고 있다. 빅데이터(big data)는 다양한 생산자와 기기(device)에 의해 빠르게 생산되어 기본적으로 양은 많고 형태와 속도도 다양한 데이터이다. 위키피디아(en.wikipedia.org/wiki/Big_data)는 빅데이터를 허용된 시간에 데이터를 수집(capture), 정리(curate), 관리(manage), 처리(process)할 수 있는 통상 사용하는 소프트웨어 도구의 능력을 넘어서는 크기를 가지는 데이터군(data set)이라 정의하였다. 함유근과 채승병(2012)은 빅데이터를 보통 수집에서 수천 테라바이트 정도의 거대한 크기를 갖고 여러 가지 다양한 비정형 데이터를 포함하고 있으며 생성-유통-소비(이용)가 몇 초에서 몇 시간 단위로 일어나 기존의 방식으로 관리와 분석이 매우 어려운 데이터의 집합을 의미한다고 정의하였다. 이런 빅데이터의 정의는 데이터의 규모와 처리에 초점을 맞추고 있다.

양적 변화가 질적 변화를 수반하는 것처럼 데이터의 양이 많아지면서 데이터의 가치도 달라진다. 빅데이터는 데이터의 규모가 방대하고(volume), 데이터의 종류가 다양하며(variety), 데이터 처리 및

분석을 적시에 해결해야 하는(velocity) 특성을 가지고 있으며, 그 결과로 새로운 가치(value)를 창출해 낼 수 있어야 한다(안창원, 황승구, 2012). Beyer & Laney(2012)는 빅데이터를 향상된 의사결정, 인사이트(insight) 발견, 그리고 과정 최적화(process optimization)를 할 수 있는 처리(processing)의 새로운 형태를 요구하는 많은 양(high-volume), 빠른 속도(high-velocity), 그리고 높은 다양성(high-variety)을 갖는 정보 자산(information assets)이라고 하였다. 결국 빅데이터는 정확한 의사결정(decision making), 통찰력, 그리고 효과적인 처리 등을 통해 새로운 가치를 창출할 수 있는 데이터의 집합(set)이다.

Informatica(2012)는 [그림1]처럼 빅데이터의 구성도를 그리고 빅데이터를 빅거래데이터(Big transaction data), 빅상호작용데이터(Big interaction data), 그리고 빅데이터처리(Big data processing) 등의 세 가지 주요한 요소의 통합이라 정의하였다. 데이터의 생산 과정과 처리에 중점을 두는 Informatica의 개념 정의에 의하면 데이터의 종류는 빅거래데이터와 빅상호작용데이터로 구분할 수 있다.

빅데이터는 표본 데이터는 많은 변수가 있고 정규분포를 가정하는 확률적 모형을 이용해 통계학적 예측을 하고 있지만 비선형적 상호작용 속에서 발견되는 패턴을 중시하여 정규분포를 가정하기 보다는 멱함수(power law) 곡선을 따르는 척도 없는 네트워크(scale-free network) 모형에 의한 징후 발견과 네트워크적 예측을 한다. 척도 없는 네트워크(scale-free network)는 노드 각각의 연결 수가 평균을 중심으로 양극단으로 갈수록 줄어드는 것이 아니라 양극단에 최고와 최저의 연결 수를 가진 노드가 상당히 존재한다(Barabasi & Alber, 1999). 척도 없는 네트워크에서는 허브(hub)와 매개자(broker)가 발견되며 허브는 네트워크 내에서 정보 수집과 확산에 상당한 영향력을 가지고 있고 매개자는 집단과 집단의 정보 다리 역할을 한다.



[그림1] 빅데이터의 구성도

1.2. 빅데이터의 활용

빅데이터는 대용량의 데이터를 저장, 수집, 발굴, 분석, 비즈니스화 하는 일련의 과정(유민호, 2012)으로 현황 파악 및 미래 예측을 할 수 있어 현재 공공, 의료, 마케팅, 정보통신 등 다양한 분야에서 도입기 단계의 활용을 하고 있다.

공공분야는 국가적 차원의 방대한 데이터를 바탕으로 이루어지는 수자원 관리, 스마트 그리드, 재난 방재 영역 등이 대표적이다(안창원, 황승구, 2012). 공공분야의 빅데이터 활용은 대부분 열린 정부(open government)와 밀접한 관련성이 있어 정부의 투명성과 개방성을 높이고 수준 높은 분석을 통해 공공분야의 경쟁력 강화할 수 있다. 미국 샌프란시스코는 과거 8년간의 범죄 데이터 분석을 통해 새로운 범죄가 일어날 가능성이 높은 지역을 예측하고 경찰을 전략적으로 사전 배치하여 실제 강력 범죄 10개 중 7개를 미연에 방지하여 범죄 억제 효과를 보고 있다(김승윤, 2012).

의료분야는 각종 임상 및 검진 기기(device) 데이터의 통합, 병원 간 검사 기록과 연구 데이터 공유를 통해 병의 징후를 조기에 발견하고 질병을 예측하는데 빅데이터 도입과 활용이 확대되는 추세이다. 미국의 경우 빅데이터와 관련하여 가장 주목 받는 영역 중 하나가 의료 분야인데 미국의 의료개

혁과 관련되어 의료기관, 환자, 정부, 의료보험회사를 하나로 통합하는 'Health 2.0' 프로젝트에서 빅데이터가 큰 역할을 담당할 것으로 여겨진다. 미국 Seton Health Care Family사는 연간 200만 명에 달하는 환자들의 복잡한 진료 정보를 IBM의 왓슨 시스템을 도입하여 환자가 미래에 겪을 수 있는 질환이나 증상을 미리 예측하여 이를 예방하는 통합되고 확장된 의료 서비스를 제공할 계획이다(안창원, 황승구, 2012).

마케팅분야는 소셜비즈니스인텔리전스(social business intelligence, Social BI)를 통해 빅데이터를 주로 활용하고 있다. Social BI는 웹과 소셜미디어의 실시간 모니터링으로 제품 및 소비자 욕구와 관련 데이터들을 통해 의미를 찾아내고 지속적으로 통찰력(insight)을 도출하여 효율적인 의사결정을 지원하는 프로그램이다(Hinchcliffe, 2011). Social BI는 시장의 정적, 동적 요인을 보다 정밀하게 분석하여 이에 따른 적시 수요 예측 및 선제적 경영지원에 초점을 두어 효과적인 의사결정을 돕고 있다.

정보통신분야의 빅데이터 활용은 스마트폰의 대중화와 소셜의 생활화로 위치기반서비스(LBS)의 텍스트, 이미지, 동영상 등의 다양한 형태의 개인 데이터들과 밀접한 관련이 있다. 빅데이터를 기반으로 한 기업의 타겟 마케팅커뮤니케이션과 실시간

소비자 분석 및 적기(just in time) 개인화 서비스를 제공하기 위한 솔루션 개발과 플랫폼 구축이 확대되고 있다. 현재 개인의 위치 정보 수준에서 진행되고 있는 서비스들은 점차 사용자의 행동 패턴, 이력, 선호도, 주변 상황에 따라 적절한 지식을 제공하는 서비스 형태로 점차 진화할 것으로 예상된다(안창원, 황승구, 2012). 특히 소셜미디어에서 발생하는 각종 데이터들 속에서 소비자의 반응과 개인 프로파일이 결합된 패턴을 추출하여 고객계층을 세분화시켜 기업들은 타겟 마케팅을 실현할 수 있다(김상락 외, 2012).

빅데이터의 활용은 시스템의 구축보다는 연구의 목적을 정하고 수많은 데이터 속에서 목적에 부합하는 데이터를 찾아내고 어떻게 처리할 것인지를 분명하게 하는 서비스가 중요하다. 빅데이터 서비스는 분석을 위해 수많은 데이터를 수집, 축적하여 빅데이터 속에서 패턴을 추출하고 징후를 발견하고 흐름을 읽고 영향력을 파악하기 위해 분석 결과의 시각화(visualization)에 대한 통합적 사고와 해석 능력이 매우 중요하다.

2. 빅데이터 분석(Big data analysis)

2.1. 텍스트마이닝과 시맨틱네트워크분석

매일 엄청난 양의 데이터들이 인터넷에 쌓이고 있고 과학기술의 발달로 이런 데이터들을 분석할 수 있는 프로그램들이 개발되면서 빅데이터 분석이 본격화 되고 있다. 빅데이터 분석은 방대한 데이터를 연관성에 따라 묶고 강도와 중심성을 계산하고 그리고 패턴을 조사하여 의미 있는 정보로 재탄생시켜 무질서한 흐름 속에서 숨겨진 의미 즉 지식을 찾아낸다. 빅데이터 분석에는 텍스트마이닝(text mining), 평판분석(opinion mining), 소셜네트워크 분석(social network analytics), 클러스터분석(cluster analysis) 등이 있다(김정숙, 2012). 텍스트 마이닝은 비정형의 텍스트데이터를 자연어처리와 형태소분석 기술에 기반하여 정제하고 유용한 단어

를 추출해 빈도 수를 보여주는 기술이다. 텍스트마이닝은 빅데이터에서 사용자가 관심을 가지는 정보를 키워드 수준이 아니라 맥락(context) 수준의 의미를 찾아내는 프로세스를 의미한다(정근하, 2010). 텍스트마이닝은 여러 문서에서 동시에 출현하는 단어의 빈도를 통해 단어의 중요도를 알 수 있다. 검색 결과의 순위나 문서나 인식의 유사성을 구하는 용도로 많이 사용되고 있다. 평판분석은 텍스트의 의미망을 통한 감성분석(sentiment analysis)으로 웹과 소셜미디어 상의 텍스트를 실시간으로 수집해 긍정, 부정, 중립의 여론을 판별하는 기술이다. 클러스터분석은 비슷한 특성을 가진 개체들 끼리 묶어서 유사한 속성을 가진 그룹을 발굴하는 데 사용한다. 마케팅을 위한 목표집단의 분류에 도움이 된다(김정숙, 2012). 소셜네트워크분석은 수학의 그래프이론에 근거하여 소셜네트워크 연결구조 및 연결강도 등을 바탕으로 이용자의 영향력을 측정하고 허브 역할을 하는 이용자를 찾는 데 활용한다. 소셜네트워크분석은 사람, 조직, 기관들의 관계를 다루는 에고네트워크분석(ego network analysis)과 단어 또는 개념들의 관계를 분석하는 시맨틱네트워크분석(semantic network analysis)이 있다.

텍스트들의 연결 관계와 패턴을 시각화해 의미를 찾을 때는 빅데이터 분석 중 소셜네트워크분석의 한 종류인 시맨틱네트워크분석을 주로 활용한다. 빅데이터는 비정형인 텍스트가 서로 연결되어 있는 구조이기에 구성 요소들 사이의 관계를 찾아내어 구조를 파악하는(Wasserman & Faust, 1994) 소셜네트워크분석을 텍스트에 적용한 시맨틱네트워크분석이 적합하다. Wang & Rada(1998)은 시맨틱네트워크분석을 소셜네트워크를 기반으로 개념을 노드로 나타내고 개념간의 관계를 연결로 나타낸 그래프로 정의한다. 개념은 단어나 구로 표현되는 정보 단위이며 의미는 다른 개념들과의 관계 속에 내재되어 있는 것이고 관계는 개념들 간의 연결을 나타내는 개념의 특정 범주를 의미한다(Khalifa & Liu, 2006). 시맨틱네트워크는 다양한 개념들을 연결하

고 관계 짓는 것이다(심홍진 외, 2011). 개념(concept)은 관련된 단어들의 합성체로서 사회네트워크에서의 노드와 같고 개념 간 연결은 서술(statement)이며 네트워크분석의 선이다. 핵심단어들이 결합되는 형태에 따라 특정한 의미가 발생되며 동시 발생빈도는 그 개념들의 관계를 형성한다(박한우 외, 2004).

시맨틱네트워크분석은 행위자 간 연결성을 중시하는 소셜네트워크분석과 달리 단어들의 공유된 의미를 토대로 체계적 구조를 분석하는데 주안점을 두고 있다(Doerfel & Barneff, 1999). 시맨틱네트워크분석은 핵심(prominent)단어 간의 의미론적 연관(semantic association)(Danowski, 1993)이 중요한 요소이고 핵심단어의 동시발생 빈도는 소셜네트워크 관점에서 중요한 관계(정성훈, 2012)가 된다. 시맨틱네트워크분석을 통한 빅데이터 분석의 장점은 표준화되지 않은 텍스트 자료로부터 구조화된 형태의 정보를 추출함으로써 커뮤니케이션 과정의 패턴과 의미를 시각화할 수 있다는 점이다.

빅데이터 분석을 통해 핵심단어의 빈도와 단어 간 관계를 분석함으로써 전체 데이터에 대한 구조를 시각적으로 나타내고 의미체계를 파악할 수 있다. 또한 기업의 의미망을 이용해 기업의 스키마를 만들 수도 있다. 시맨틱네트워크분석을 통해 의미(semantic) 간 네트워크를 분석하고 가시화하며 노드수, 링크수, 밀도, 연결정도중심성의 4가지 지표를 이용하여 영향력을 파악할 수 있다. 노드수, 링크수, 밀도는 네트워크의 전체적인 구조를 분석하기 위해 사용하고 연결정도중심성은 각 노드가 네트워크 구조 내에서 얼마나 중심에 위치해 있는지를 나타내는데 활용한다. 시맨틱네트워크분석은 핵심단어 간 의미론적 연결성을 확인할 수 있어 잠재수요기업을 과학적으로 제시할 수 있을 것이다.

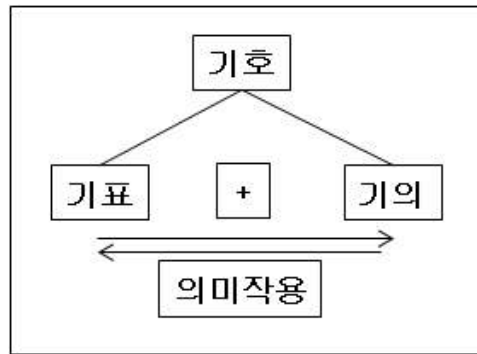
2.2. 빅데이터 분석과 기호학

텍스트의 빅데이터는 주로 단어와 단어가 연결되어 만들어진 텍스트의 집합(set)이다. 텍스트는 생각이

나 감정을 글로 표현할 때 사용되며 주어와 서술어의 체계를 갖추고 있다. 텍스트의 빅데이터 형태를 단순화시켜 보면 단어와 단어가 연결되어 있고 상호작용하면서 문장을 만들고 문장과 문장이 쌓이고 또 상호작용하면서 의미와 패턴을 만들고 있다. 결국 텍스트의 단어들은 빅데이터 분석을 위한 데이터의 원천(source)이 된다. 텍스트의 빅데이터 분석은 데이터의 원천인 단어를 기반으로 단어의 빈도를 계산하고 단어와 단어의 연결 형태를 파악하여 의미체계를 추론한다. 특히 단어의 연결 형태를 통해 의미와 연결망을 파악하는 빅데이터의 분석은 시맨틱네트워크분석이다. 시맨틱네트워크는 개념을 노드로 나타내고 개념 간의 관계를 링크로 나타낸 그래프(Wang & Rada, 1998)로 특정 개념을 정의하고 그 개념과 연관되어 있는 단어들로 개념 연상(association)의 네트워크를 그려 개념을 이해하는데 도움을 준다. 개념은 단어나 구로 표현되는 정보단위이며 의미는 다른 개념들과의 관계 속에 내재되어 있는 것이다. 관계는 개념들 간의 연결을 나타내는 개념의 특정 범주를 의미한다(Khalifa & Liu, 2006). 개념인 단어의 연결 관계를 통해 의미를 파악하는 지식은 기호학과 밀접한 관련이 있다.

김광철 외(2004)는 기호학을 개별적인 기표들이 특정한 어법에서 다른 기표들과 맺는 상관관계를 통해 의미를 창출하는 방법을 연구하는 학문이라고 정의하였다. 오정근(2010)은 기호학(semiotics)을 상징체가 어떤 구조로 만들어져 있으며 어떤 의미를 품고 있는가를 분석하는 학문이라고 정의했다. 상징체인 텍스트의 상관관계를 분석하면 기표들의 구조와 의미를 파악할 수 있다. 기호학은 기호와 의미가 만들어지는 구조와 형태를 분석하는 학문으로 시맨틱네트워크분석을 통해 단어를 상관관계를 시각화한 의미망과 유사하다.

Barthes가 정교화한 [그림2]와 같은 기호의 삼부 모형(semiotic triangle)을 살펴보면, 기호(sign)는 기표, 기의, 그리고 기호 자체의 세 가지 기본요소로 이루어지고 기호 자체는 기표와 기의가 연합하



[그림2] 기호의 삼부모형

여 만들어 낸 새로운 요소이다(김경용, 1994 재인용). 기호(sign)는 기표(signifier)와 기의(signified)의 두 요소가 결합하여 의미작용(signification)을 하면서 만들어지는 것으로서 기표는 의미의 표현체로 무엇을 표현하기 위한 단어로 형식을 말하고 기의는 기호가 대변하는 추상적 개념으로 의미있는 내용을 말한다.

김경용(1994)은 의미작용을 기표와 기의를 결합시키는 작용이고 기호를 만들어 낼 때와 기호의 의미를 풀이할 때에도 일어난다고 하였다. 의미작용은 디노테이션(denotation)과 코노테이션(connotation)의 두 개의 층의가 작용하고 있다. 코노테이션(connotation)은 기의로 어떤 단어가 의미하는 주관적 문화적 감정적 연상이고 디노테이션(denotation)은 기표로 표면적이고 글자적 의미이다.

Barthes는 1차 기호(the primary sign)는 외시의미(denotation)로 객관적 의미의 수준을 나타냈고 또 1차 기호의 표현과 내용이 합쳐져서 2차 기호의 새로운 표현(기표)과 내용(기의)이 되어 주관적 의미의 수준, 즉 개인이나 문화에 따라 다른 의미를 부여할 수 있는 자의성의 차원으로 공시(함축)의미(connotation)을 가지게 된다고 설명하였는데 이는 텍스트를 분석하는데 유용한 틀을 제공한다(박영원, 2003).

Greimas의 행위소모형(actantial model)은 주체(subject)와 대상(object)의 대립쌍으로 구성되어 있고 의소(eme, 의미의 최소 단위)를 세분하여 담론

의 심층구조를 살피고 이를 통해 표층의 의미를 읽고 다시 텍스트로 표출되는 의미들을 구조적으로 분석한다(오정근, 2010).

텍스트 빅데이터를 기호학의 관점에서 살펴보면 단어는 의소이고 단어와 단어의 네트워크인 텍스트는 의미를 만들고 있는 삼부모형의 확장으로 시맨틱네트워크다. 시맨틱네트워크는 다양한 단어들의 복잡하고 다차원적인 연결과 관계의 시각화이다. 단어와 단어의 상관관계는 사람들이 단어를 이해하고 공유하고 기억하는 방식을 알 수 있는 의미론적 지식구조도 알 수 있게 한다. 빅데이터의 시맨틱네트워크분석을 실시하면 의미망의 시각화를 통해 기호학의 모형을 구성하는 주체와 대상의 단어를 찾고 단어의 상관관계로 담론의 심층구조를 파악할 수 있다. 의미망은 주체(기의) 단어와 대상(기표) 단어의 연결 상태와 단어별 중요성을 보여주면서 한 단어가 다른 단어를 설명 또는 예측해 주는 역할을 알 수 있게 한다.

기술마케팅의 잠재 수요기업은 기술이전을 통해 상용화를 창출할 개연성이 높은 기업이다. 기호학의 관점에서 보면 기술의 잠재 수요기업은 특정기술과 관련된 많은 단어가 강하고 밀접하게 연결되어 있을 것이다. 역으로 특정기술과 강하고 밀접하게 연결되어 있는 기업은 기호학의 모형에 의해 잠재 수요기업일 것이다. 빅데이터의 시맨틱네트워크 분석의 결과는 의미망의 시각화를 통해 특정기술과 강하고 밀접하게 많이 연결되어 있는 잠재 수요기

업을 단순명료하게 보여준다. 네트워크의 시각화는 텍스트를 구조적이고 체계적으로 보여주기 때문에 단어와 단어의 상관관계와 계층적 구조를 보여 주는 원리를 이용하여 기술관련 분석단어와 핵심단어를 통해 잠재 수요기업을 찾게 해준다.

III. 연구 방법

1. 분석모형 및 연구대상

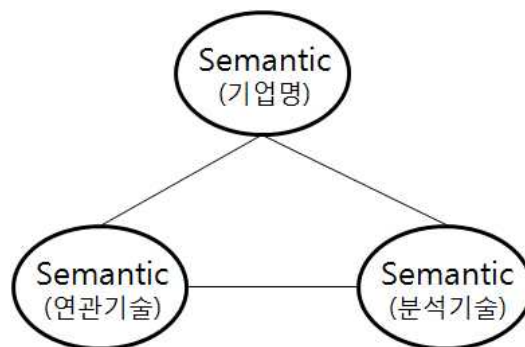
빅데이터 분석은 방대한 데이터 속에서 네트워크 강도와 중심성의 시각화(visualization)를 통해 관계와 패턴을 확인하여 숨겨진 의미를 밝혀내고 예측한다. 빅데이터 분석은 서로 연관되어 있는 노드들을 통해 정보와 가치를 찾아 내는 방법이다. 빅데이터는 여러 분석들 중에서 텍스트마이닝과 시맨틱네트워크분석을 주로 사용하고 있다. 빅데이터에서 단어 수준이 아니라 맥락(context)의 수준에서 주요 정보를 찾는 분석에는 텍스트마이닝(textmining)이 적합하다. 개념 간의 관계를 통해 의미를 찾아내고 분석단어와 핵심단어의 동시 연결단어를 찾는 데는 시맨틱네트워크분석(semantic network analysis)이 적합하다. 시맨틱네트워크분석은 개념의 연결 형태를 통해 네트워크 형태의 지식 덩어리인 스키마(schema)와 기본개념에 설명적 단어들이 상층단계, 기본단계, 하부단계의 위계적 순서로 연결되어 있는 의미론적 조직(semantic

organization)을 잘 보여준다.

빅데이터 분석의 모형은 기호학의 개념에 기반하여 사람들이 특정단어를 언급할 때는 그와 연계된 의미나 개념, 유사한 맥락을 가진 단어들을 함께 언급할 가능성이 높다는 전제를 한다. 가령 사람들이 ‘인터넷’을 논의하는 중에 ‘검색’이라는 단어를 자주 언급하는 식이다. 단어의 공동출현(co-occurrence) 현상이 발생하는 논리이다(심홍진 외, 2011). 단어의 공동출현은 기호학의 ‘기호의 삼부모형’과 ‘행위소모형’으로 설명될 수 있다.

기술마케팅은 기술 사업화를 위한 기술의 이전이 중심이지만 마케팅의 효율성을 높이기 위해서는 기술의 특성과 제한된 시장으로 인해 잠재 수요기업을 찾는 것이 매우 중요하다. 기술마케팅에서 잠재 수요기업 발굴이 중요해진 가운데 새로운 자료와 분석 방법을 이용하는 빅데이터 분석의 유용성에 주목하게 된다. 빅데이터를 활용한 기술마케팅의 잠재 수요기업 발굴은 빅데이터 분석을 통해 의미 있는 단어들의 연결 상태와 중심성으로 기술을 필요로 하는 기업을 찾아 내는 방식이다. 또한 빅데이터 분석은 기업 스키마를 통해 전략적인 마케팅 커뮤니케이션을 실행할 수 있는 정보들도 확인할 수 있다.

기호학적 모형과 단어의 공동출현 시각화하는 시맨틱네트워크를 토대로 하여 기술마케팅의 잠재 수요자 발굴을 위한 빅데이터 분석의 모형을 그려보면 [그림3]과 같다.



[그림3] 분석모형

분석기술의 시맨틱네트워크는 주체와 대상의 기술구조로 연결된 연관기술과 기업명이 어떻게 연결되어 있고 무엇에 관한 것인가를 밝혀 준다. 수집된 빅데이터를 정제하여 시맨틱네트워크분석을 실시하면, 분석기술, 연관기술, 기업명 등이 공동출현(co-occurrence) 하는 네트워크의 시각화가 이루어진 분석모형이 만들어 진다. 분석모형은 기호학적 모형처럼 분석기술을 중심으로 기술관련 주요 단어가 연결되어 있고 이 기술관련 주요 단어들과 기업명이 공동출현으로 연결되어 있다. 기업명이 공동출현으로 연결되어 있다는 것은 분석기술과 상관관계가 있고 연결정도도 높을 수 있다. 이 연구에서는 잠재 수요기업을 분석기술 및 연관기술들과 공동출현으로 연결되어 있는 기업명으로 조작적 정의를 하였다. 분석모형을 바탕으로 연구문제를 정해 보면 다음과 같다.

연구문제1] 분석기술 빅데이터의 분석을 통해 확인된 주요기술들은 무엇인가?

연구문제2] 분석기술 빅데이터의 분석을 통해 발굴된 잠재 수요기업은 어디인가?

연구문제3] 잠재 수요기업과 연관되어 있는 설명 단어들은 무엇인가?

2. 자료수집 방법

기술마케팅의 잠재 수요기업 발굴 방법으로 빅데이터를 활용하기 위해 한국표준과학연구원(KRISS)의 신기술을 대상으로 선정하였다. KRISS는 약 250개의 이전가능 기술을 보유하고 있으며 연간 약 30건의 기술이전 계약을 체결한다. KRISS 기술사업화센터는 기술마케팅을 위해 페이스북마케팅을 활발히 실행 하는 등 새로운 기술마케팅의 방안을 모색하고 있는 기술이전기관(TLO)이다. KRISS의 기술은 기술사업화가 가능한 다양한 특허와 기술이 있고 새로운 기술마케팅에 대한 관심이 높은 TLO가 있기 때문에 본 연구에 적합하다고 판단되어 연구 대상으로 선정했다.

연구의 목적을 위해 KRISS가 보유하고 있는 ‘배열검출기형 분광복사계 성능평가 기술’과 ‘초음파를 이용한 진공압력 측정센서’를 연구 대상 기술로 선정하였다.

자료수집 방법은 인터넷에 분산저장 되어 있는 문서를 수집하는 크롤링(crawling)방식을 통해 기업의 공식성을 갖고 있는 웹페이지, 개인들이 운영하는 블로그, 공신력 있는 정보를 제공하는 뉴스 등에서 자료를 수집하였다. 자료수집은 데이터 수집 및 처리 프로그램인 THE SCRM¹⁾을 활용하였다.

분석단어는 기술소개서를 기초자료로 활용하여 기술의 관련 범위를 대(사업영역)>중(제품영역)>소(기술영역)로 범주화하여 선정하였다. 분석단어를 추출한 후 3명의 Coder가 협의를 통해 최종적으로 분석단어를 분광복사계, 광계측, 진공압력측정센서, 진공압력모듈 등 4개를 선정하였다.

3. 연구 방법

3.1. 텍스트마이닝

단어 빈도와 시맨틱네트워크분석을 위한 매트릭스 데이터를 만들기 위해 텍스트마이닝을 실시하였다. 수집된 데이터에서 분석단어와 관련 없는 데이터를 삭제하고, 어근과 어미를 분리하고 특수문자를 제거하는 단어의 정제화²⁾를 실시하였다. 다음으로 띄워쓰기가 되어 있으나 같은 의미로 사용되어진 단어를 통합하여 단어를 정교화 하였다.

분석을 위해 THE SCRM을 이용하여 연관있는 기술명과 기업명들만 도출하여 빈도를 계산하였다.

- 1) THE SCRM는 크롤링 방식을 이용해 Web상의 데이터를 서비스별로 자동 수집하여 처리해주는 데이터웨어가 개발한 데이터 처리 솔루션이다. 전통적인 조사의 한계를 극복하고 인터넷의 데이터를 자동으로 수집할 수 있고 수집된 데이터를 정제한 후 결과를 정리 분석할 수 있는(박한우, 2012) 조사도구의 필요성 때문에 THE SCRM은 개발되었다.
- 2) 정제화 작업에 있어서 연구자의 주관적 판단에 의한 분류작업이 이루어지지 않도록 최대한 주의 기울였다.

시맨틱네트워크분석을 위해 (단어)×(단어)의 1모드 동시출현 빈도 값인 공동출현빈도(co-occurrence) 대칭형 매트릭스자료를 만들었다.

3.2. 시맨틱네트워크분석

단어들 간의 의미망으로 빅데이터에 내재되어 있는 의미를 파악하고 분석기술의 관계를 시각화하여 공동출현 관계에 있는 기업명을 통해 잠재 수요기업을 발굴하기 위해 시맨틱네트워크분석을 실시하였다. 시맨틱네트워크분석은 Ucinet6³⁾를 사용하였다. Ucinet6를 사용해 단어들 간 연결 구조를 파악하고 연결정도중심성을 분석하고 관계의 정도를 계량화하였다. 다음으로 Ucinet6의 NetDraw⁴⁾ 프로그램을 이용하여 단어들 사이의 관계를 명료하게 표현하고 연관단어들 사이의 네트워크를 시각화하였다. 단어들의 각 빈도 수를 노드 크기에 반영하여 데이터 상에서 언급된 정도를 보여주고, Spring Embedding과 Ego Network를 이용한 레이아웃으로 연결구조를 가장 명료하게 하였다.

3.3. 잠재 수요기업 정의 및 발굴 방법

잠재 수요기업은 기술의 이전을 통해 사업화할 가능성이 높은 기업으로 분석기술과 연관성이 높고, 기술의 활용 관련성이 높은 분야에서 현재 사업을 하고 있는 기업을 말한다. 빅데이터의 텍스트마이닝과 시맨틱네트워크분석을 통해 잠재 수요기업을 발굴하기 위해 잠재 수요기업을 조작화 하였다. 잠재 수요기업 대해서는 현재까지 일반적으로 받아들여지는 기준이 존재하지는 않으나 분석기술과 반드시 연결되어 있으면서 공동으로 2개 이상의 의미있

는 기술관련 단어와 연결된 최소 빈도수 2 이상인 기업으로 정의한다. 핵심 단어는 분석단어 외에 시맨틱네트워크분석을 통해 연결중심성 및 빈도 수 기준으로 도출한 주요단어 20개이다(서일원 외, 2013). 만일 분석단어가 복합단어인 경우 띄워쓰기로 분리되어 있거나 개별단어가 단독으로 사용되어 있는 경우 분석단어의 개별단어가 모두 연결되어 있어야만 분석단어와 같이 인정될 수 있다.

잠재 수요기업 발굴 방법은 먼저, 빅데이터의 전체 기술관련 단어들 중에서 각 데이터별 분석단어와 기업명을 제외하고 연결중심성과 빈도가 높은 상위 20개의 핵심단어를 선정하였다. 두 번째는 20개의 핵심단어를 이용하여 빅데이터의 시맨틱네트워크분석을 실시하여 1차적으로 네트워크중심도와 발굴된 기업의 연결정도중심성(degree centrality)을 통해 잠재 수요기업을 발굴하였다. 기업의 연결정도중심성은 기업노드 한 점에 연결된 다른 점의 수의 합으로 측정되며, 연결된 점의 많고 적음의 여부가 그 기준이 되어 그 점이 중심에 위치한 정도를 본다(손동원, 2002). 네트워크 시각화에서 공동출현을 보이고 기술관련 주요단어가 많이 연결된 기업일수록 연결정도중심성 수치가 높고, 잠재 수요기업의 가능성이 높다.

잠재 수요기업 발굴 방법의 단계는 첫 번째, 분석단어 전체 네트워크의 그림을 그린 후 핵심단어의 에고네트워크를 그린다. 두 번째, 각 핵심단어의 에고네트워크에서 발굴된 기업명을 찾아 리스트를 작성하여 분석단어와 2개 이상의 핵심단어들과 공동출현한 기업을 찾는다. 세 번째, 단어와 기업명 사이의 코사인 거리 유사계측 값을 확인하고, 마지막으로 선의 굵기를 통한 연결강도를 통해 잠재수요기업을 발굴한다. 잠재 수요기업 발굴 과정에서 공동출현, 코사인 유사계측, 연결강도를 기업 발굴의 기준값으로 사용하였다.

권선필(2008)에 의하면 동시출현(Collocation 또는 Co-occurrence)은 두 개의 단어가 한 문장이나 텍스트 안에서 같이 발생하는 현상으로 동시출현한

3) Ucinet은 Freeman이 개발한 다양한 연결망 분석기법을 활용할 수 있는 종합적인 프로그램. 응집력(Cohesion), 하위집단(Subgroup), Ego Network, 중앙성(Centrality), 중앙과 변방(Core/Periphery), 역할/위치(Roles & Positions) 등을 분석한다(Borgatti, Everett, & Freeman, 2002).

4) 넷드로우(NetDraw): Ucinet의 시각화 프로그램. 노드와 연결 관계를 그래프 형태로 그릴 수 있다.

두 단어는 서로 밀접한 관계가 있다고 할 수 있다. 기술관련 핵심단어와 공동연결이 많은 기업일수록 잠재 수요기업일 가능성이 높다고 보았다. 코사인 유사계측을 통해 기술관련 주요단어와 연결된 기업의 거리를 살펴보았다. 코사인 유사계측은 (단어)×(단어)인 1모드 코사인 대칭형 매트릭스로 코사인 행렬에서 수치는 단어 간 거리를 의미한다(박한우 외, 2004). 분석단어와 기업명이 1에 가까운 수치일수록 잠재수요기업일 가능성이 높다. 연결강도는 네트워크 그림에서 선의 굵기로 단어와 기업명 연결의 강도(the strength of ties)를 의미한다. 노드 간의 관계에 있어 강한 연결과 약한 연결의 비교는 관계의 친근성(closeness of a relationship)(손동원, 2002)을 이용한다. 기술관련 핵심단어와 기업명 사이의 선의 연결 강도가 강할수록 공동출현 빈도가 많은 것으로 분석기술과 관련 있는 잠재 수요기업일 가능성이 높다.

2차적으로 홈페이지를 통해 재검증하는 과정을 거쳤다. 발굴한 잠재 수요기업의 재검증 과정을 통해 잠재 수요기업의 적확성(definitude)을 높였다. 재검증 과정에서는 기업의 연결정도중심성(degree centrality)과 기업 홈페이지를 통해 사업 분야를 확인하였다.

IV. 연구 결과

1. 기술 핵심단어 선정

각 신기술의 분석단어로 데이터를 수집한 결과, 데이터는 분석단어별로 최소 109개에서 최대 137,041

개이다. 하지만 현재 수집 가능한 최대 데이터는 3,000개인 관계로 전체 데이터가 아닌 최대 3,000개의 데이터가 실제로 수집되었다. 현재 결과 페이지에서 정확도 순으로 데이터를 가져오기 때문에 분석단어에 적합한 데이터는 대부분 수집되었다고 할 수 있다. 분석단어별 데이터 수집의 결과는 [표1]와 같다.

신기술별로 살펴보면, <배열검출기형 분광복사계 성능평가 기술>관련한 분석단어는 <분광복사계> 225개와 <광계측> 3,000개로 총 3,225개의 데이터가 수집하였다. <초음파를 이용한 진공압력 측정센서 및 모듈>의 분석단어는 <진공압력측정센서> 690개와 <진공압력모듈> 1,224개로 총 1,914개를 수집하였다.

빅데이터 분석을 통해 확인된 핵심단어들은 연결정도중심성을 기준으로 연결정도 수치가 높은 단어에 빈도수를 참고하여 분석단어별로 상위 20개씩 선정하였다. <배열검출기형 분광복사계 성능평가 기술>의 선정된 주요단어는 [표2]과 같다.

분석단어인 ‘분광복사계’의 분석단어와 기업명을 제외하고 첫 번째 핵심단어는 연결정도 78, 연결정도중심성 0.04, 빈도 7의 값을 보인 ‘태양광’이 선정되었다. 이어서 ‘근적외선’이 연결정도 78, 연결정도중심성 0.04, 빈도 7로 두 번째로 높게 나타났고 연결정도가 66이고, 연결정도중심성이 0.034이며, 빈도가 16인 ‘흑체’와 연결정도 54와 연결정도중심성 0.028, 빈도 18의 ‘광조도계’도 핵심단어로 선정되었다. ‘휘도’는 전체 데이터에서 두 번밖에 언급 안 되었지만, 연결정도가 48이고 연결정도중심성이 0.025를 보이고 있어 주요단어로 선정되었다.

[표1]데이터 수집 결과

신기술명	분석단어	활용 데이터
배열 검출기형 분광복사계 성능평가 기술	분광복사계	225
	광계측	3,000
초음파를 이용한 진공압력 측정센서 및 모듈	진공압력 측정센서	690
	진공압력모듈	1,224

[표2] 배열 검출기형 분광복사계 성능평가 기술 핵심단어 상위 20개

분광복사계				광계측		
번호	단어	연결정도 (중심성)	빈도	단어	연결정도 (중심성)	빈도
1	태양광	78(0.040)	15	광계측장비	1009(0.055)	228
2	근적외선	78(0.040)	7	레이저	815(0.044)	149
3	흑체	66(0.034)	16	LED	710(0.039)	82
4	광조도계	54(0.028)	18	광통신	697(0.038)	104
5	휘도	48(0.025)	2	측정	626(0.034)	73
6	Spectroradiometer	45(0.023)	18	나노	390(0.021)	26
7	고해상도	43(0.022)	7	모듈LED	345(0.019)	7
8	분광계	43(0.022)	7	반도체	343(0.019)	57
9	카메라	41(0.021)	2	LED광계측기기	319(0.017)	55
10	절대온도	39(0.020)	11	조도측정장비	318(0.017)	30
11	광휘도	38(0.019)	1	광원	305(0.017)	43
12	광휘도계	32(0.016)	4	광계측센서	293(0.016)	9
13	복사조도	28(0.014)	2	일반조명기구	287(0.016)	25
14	의료영상	27(0.014)	1	광계측센서기기	272(0.015)	39
15	감지계	25(0.013)	3	계측제어기기	219(0.012)	6
16	에어로졸	25(0.013)	3	광계측제어기기	204(0.011)	46
17	고도계	25(0.013)	3	분광	203(0.011)	3
18	광도센터	24(0.012)	2	광섬유	203(0.011)	40
19	태양전지	24(0.012)	8	광측정	199(0.011)	14
20	광측정	23(0.012)	5	정밀광계측	190(0.010)	26

두 번째 분석단어인 ‘광계측’과 관련하여 분석단어와 기업명을 제외하고 ‘광계측장비’가 연결정도 1,009, 연결정도중심성 0.055, 빈도 228로 가장 높게 나타나 첫 번째 주요단어로 선정되었다. 이어서 ‘레이저’와 ‘LED’의 연결정도(연결정도중심성)가 각각 815(0.044)와 710(0.039)으로 높게 나타났다. 연결정도가 697이고 연결정도중심성이 0.038이며 빈도 104의 ‘광통신’과 연결정도 626, 연결정도중심성 0.034, 빈도 73의 ‘측정’도 핵심단어로 선정하였다.

초음파를 이용한 진공압력 측정센서 및 모듈의

핵심단어는 [표3]과 같다.

‘진공압력 측정센서’ 분석단어로 수집한 데이터에서는 분석단어를 제외하고, ‘압력센서’가 연결정도 1,557, 연결정도중심성 0.052, 빈도 140으로 가장 높게 나타나 첫 번째 핵심단어로 선정되었다. 이어서 ‘온도’가 연결정도 979, 연결정도중심성 0.033, 빈도 66로 두 번째로 높게 나타났다. 그 밖에 연결정도가 609이고 연결정도중심성이 0.02, 빈도 19인 ‘장비’, 연결정도 606, 연결정도중심성 0.02, 빈도 62의 ‘압력측정’, 연결정도 531, 연결정도중심성 0.018, 빈도

[표3] 초음파를 이용한 진공압력 측정센서 및 모듈 핵심단어 상위 20개

진공압력 측정센서				진공압력모듈		
번호	단어	연결정도 (중심성)	빈도	단어	연결정도 (중심성)	빈도
1	압력센서	1557(0.052)	140	센서	1732(0.043)	82
2	온도	979(0.033)	66	측정	1628(0.041)	52
3	장비	609(0.02)	19	밸브	1530(0.038)	93
4	압력측정	606(0.02)	62	펌프	1163(0.029)	52
5	유량	531(0.018)	12	제어	1116(0.028)	53
6	게이지	480(0.016)	24	게이지	1060(0.027)	27
7	펌프	395(0.013)	12	기술	875(0.022)	77
8	진공압	357(0.012)	7	온도	717(0.018)	90
9	엔진	344(0.011)	33	압력진공	615(0.015)	19
10	진공압력	280(0.009)	31	진공펌프	597(0.015)	65
11	밸브	267(0.009)	22	압력모듈	584(0.015)	19
12	맵센서	262(0.009)	30	압력센서	572(0.014)	71
13	온도센서	261(0.009)	28	압력진공측정 압력모듈	489(0.012)	36
14	측정센서	260(0.009)	30	태양열	457(0.011)	53
15	진공펌프	253(0.008)	22	에너지	361(0.009)	27
16	탄소나노튜브	245(0.008)	19	압력게이지	354(0.009)	19
17	측정기	241(0.008)	35	진공압력	336(0.008)	36
18	반도체	239(0.008)	15	반도체	298(0.007)	29
19	진공게이지	233(0.008)	26	태양광	291(0.007)	37
20	유량계	189(0.006)	13	온수기	278(0.007)	32

12의 ‘유량’ 등을 핵심단어로 선정하였다. 다음으로 분석단어 ‘진공압력모듈’의 데이터에서 ‘진공’, ‘압력’, ‘모듈’의 분석단어를 제외하고 ‘센서’가 연결정도 1,732, 연결정도중심성 0.043, 빈도 82로 가장 높게 나타나 핵심단어로 선정하였다. 이어서 연결정도 1,628, 연결정도중심성 0.041, 빈도 52의 ‘측정’과 연결정도 1,530, 연결정도중심성 0.038, 빈도 93의 ‘밸브’, 연결정도 1,163, 연결정도중심성 0.029, 빈도 52의 ‘펌프’, 연결정도 1,116, 연결정도중심성 0.028, 빈도 53의 ‘제어’ 등을 핵심단어로 선정하였다.

2. 잠재 수요기업 발굴

시맨틱네트워크분석을 실시하여 분석단어 네트워크의 시각화를 하고 분석단어와 의미있는 핵심단어들의 공동출현 연결망을 확인하여 잠재 수요기업을 찾았다. 분석단어와 20개의 핵심단어들과 동시에 연결되어 있으면서 빈도 2 이상의 기업들을 발굴한 결과, 최종적으로 27개의 잠재 수요기업이 추출되었다. 각 분석단어별 잠재 수요기업의 추출 결과는 [표4]과 같다. <배열 검출기형 분광복사계 성능평

[표4] 잠재수요기업 추출 결과

배열 검출기형 분광복사계 성능평가 기술		진공압력모듈	
단어	잠재기업수	단어	잠재기업수
분광복사계	1	센서	4
광계측	16	측정	6

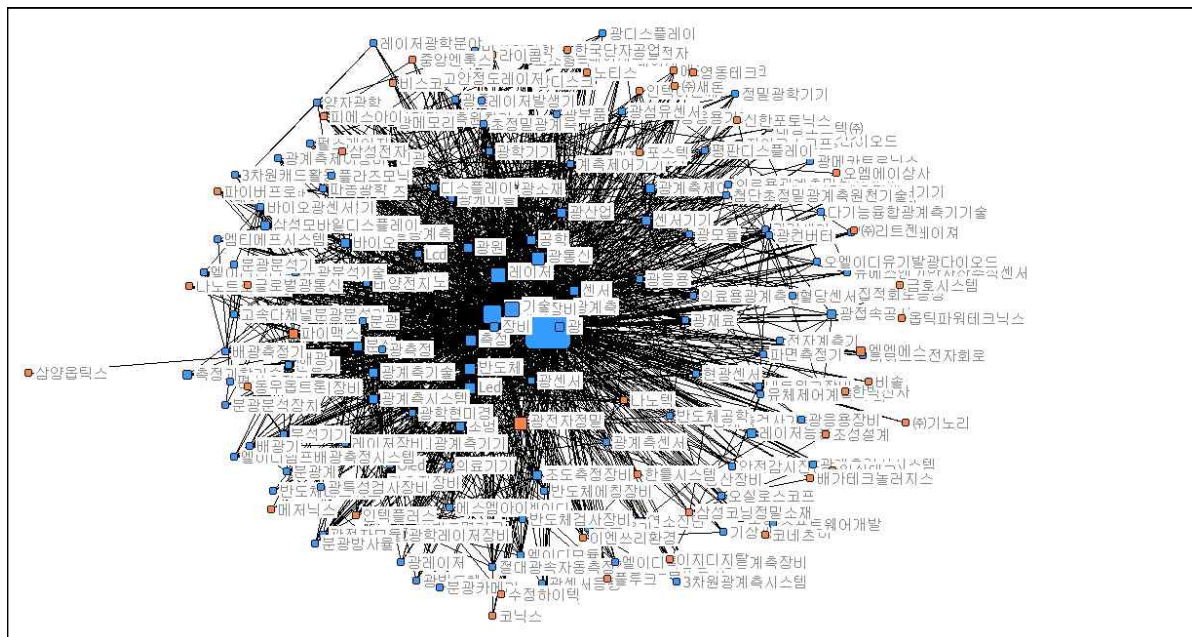
가 기술>과 관련해서는 총 17개의 기업이 추출되었고, <초음파를 이용한 진공압력 측정센서 및 모듈>관련 수요기업은 10개가 나왔다.

2.1. 배열검출기형 분광복사계 성능평가 기술 : 광계측

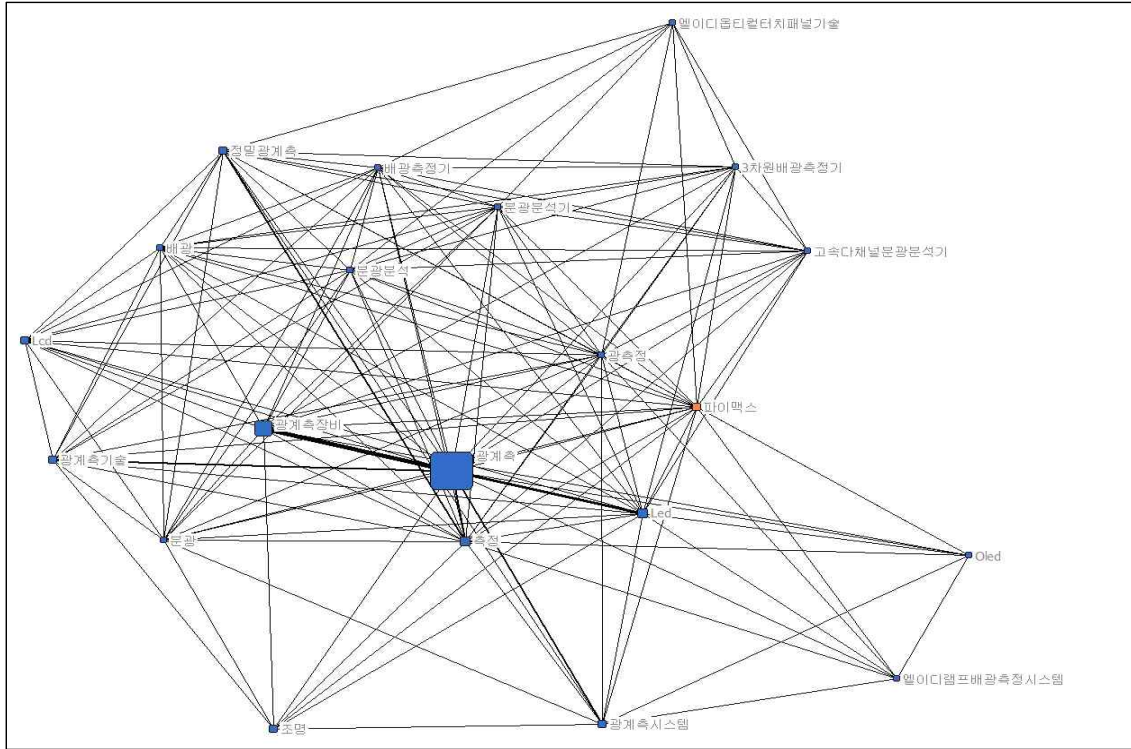
배열검출기형 분광복사계 성능평가 기술의 분석단어 ‘광계측’의 빅데이터를 활용해 잠재 수요기업 발굴해 보았다. 우선, 분석단어 ‘광계측’의 시맨틱네트워킹분석을 통해 [그림4]과 같은 전체 네트워크를 시각화하였다. [그림4]을 보면, ‘광계측’ 네트워크는 네트워크 전체의 중심화 경향을 의미하는 네트워크 중심도가 2.77%이고, 분석단어이자 언급 빈도가 가

장 높은 ‘광계측’ 노드를 중심으로 ‘기술’, ‘장비’, ‘측정’, ‘센서’, ‘레이저’, ‘광센서’, ‘반도체’ 등의 주요단어들을 볼 수 있다. 분석단어인 ‘광계측’과 20개의 핵심단어들을 통해서 총 16개의 기업을 발굴하였다.

추출된 16개의 잠재수요기업 중 하나인 ‘파이맥스’는 분석단어인 ‘광계측’과 핵심단어인 ‘광측정’, ‘광계측장비’, ‘정밀광계측’, ‘LED’, ‘조명’ 등과 공동출현으로 연결되어 있어 잠재 수요기업으로 발굴되었다. 코사인 유사계측 값을 살펴보면 ‘파이맥스’와 ‘광계측’은 0.13 코사인 값의 거리에 연결되어 있다. 주요단어인 ‘광계측장비’는 0.21, ‘정밀광계측’은 0.15, ‘LED’는 0.14, ‘조명’은 0.13, ‘광측정’은 0.23의 코사인 값이다. ‘파이맥스’는 분석단어 및 기술관련



[그림4] 분석단어 ‘광계측’ 네트워크



[그림5] 파이맥스 에고네트워크

주요단어와 동시에 연결되어 있었다.

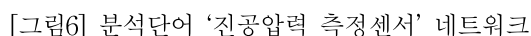
잠재 수요기업의 재검증을 위한 ‘파이맥스’의 에고네트워크분석 결과를 살펴보면, [그림5]과 같다. ‘파이맥스’의 연결정도는 179이고, 도출 과정에 있었던 ‘광계측’, ‘광측정’, ‘광계측장비’, ‘정밀광계측’, ‘LED’, ‘조명’ 등과 동시에 연결되어 있었다. 또한 ‘파이맥스’ 홈페이지를 통해 기업의 사업영역을 확인해 본 결과, 파이맥스는 정밀분석장비 전문기업으로 배광측정시스템과 적분구 등의 분야에서 사업을 하기에 배열검출기형 분광복사계 성능평가 기술 활용이 가능하고 기술이전 가능성이 있는 기업이었다.

2.2. 초음파를 이용한 진공압력 측정센서 및 모듈 : 진공압력 측정센서

초음파를 이용한 진공압력 측정센서 및 모듈의 잠재 수요기업 발굴을 위한 분석단어 ‘진공압력 측정센서’ 빅데이터의 시맨틱네트워크분석 결과는 [그

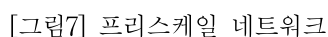
림6]과 같다. [그림6]를 보면, ‘진공압력 측정센서’ 네트워크는 네트워크중심도가 3.11%이고 분석단어 ‘진공압력 측정센서’의 세분단어인 ‘진공’, ‘압력’, ‘측정’, ‘센서’ 등이 언급 빈도가 높고 중심에 위치해 있었다. ‘진공측정’, ‘압력센서’ 등 분석단어와 관련해서 띄어쓰기된 경우 한 단어로 만들어 단어를 추출하였다. 한 문장 안에서 ‘진공’, ‘압력’, ‘측정’, ‘센서’는 동시에 사용되어 추출된 데이터이기 때문에 분석단어의 세분단어들 간에 연결이 강하게 나타난다. ‘진공’, ‘압력’, ‘측정’, ‘센서’ 등을 중심으로 ‘반도체’, ‘진공센서’, ‘압력센서’, ‘온도센서’, ‘레이저’, ‘게이지’ 등의 연결정도중심성이 높은 핵심단어들을 볼 수 있다. ‘진공압력 측정센서’로 수집한 데이터에서 최종 4개의 기업을 발굴하였다.

추출된 4개의 잠재 수요기업 중 하나인 ‘프리스케일’은 ‘진공’, ‘압력’, ‘측정’, ‘센서’ 등의 분석단어와 ‘압력센서’, ‘진공장치’, ‘압력측정’ 등과 동시에



이고 ‘진공장비’는 0.95정도의 거리로 가장 가깝게 연결되어 있었다.

잠재 수요기업의 재검증을 위한 ‘프리스케일’의 에고네트워크분석 결과는 [그림7]과 같다. [그림7]을 살펴보면, ‘프리스케일’의 연결정도는 136이고,



‘진공’, ‘압력’, ‘측정’, ‘센서’가 모두 굵은 선으로 강하게 연결되어 있어 하나의 분석단어로 간주할 조건을 만족시키고 있었다. 또한 ‘압력센서’, ‘압력측정’, ‘진공장비’ 등 2개 이상의 주요단어들과도 동시에 연결되어 있었다. ‘프리스케일’ 홈페이지를 통해 확인해 본 결과, ‘프리스케일’은 반도체 전문기업으로 초음파를 이용한 진공압력 측정센서 및 모듈의 기술 이전 가능성이 있는 기업이었다.

3. 잠재 수요기업의 스키마

잠재 수요기업들의 의미적 지식 네트워크를 통한 스키마를 추정해 보기 위해 잠재 수요기업과 연관되어 있는 설명단어들을 확인하여 보았다. 스키마는 하나의 의미적 지식이 활성화될 때 함께 떠오르는 관련된 지식들의 집합이며 기억의 덩어리로 지식의 네트워크 형태이다. 의미적 지식 네트워크는 기억 속에 저장된 지식(정보)들의 계층적 네트워크로 관련된 단어들과 이들을 서로 연결하는 선(link)으로 구성되어 있다. 단어는 장기기억 속에 저장된 의미적 지식을 말하며 연결선은 지식과 지식의 연관관계(association)이다. 특정 단어가 다른 단어들과

강한 연관관계(관련성)를 가질수록 연결고리는 짧고 굵다. 한 대상과 관련된 지식들로 구성되는 네트워크인 스키마를 알면 대상에 대한 커뮤니케이션을 보다 원활하게 실행할 수 있다.

잠재 수요기업으로 도출된 27개의 기업 중 위에서 살펴 본 2개의 잠재 수요기업과 연결되어 있는 설명단어의 네트워크를 통해 기업의 스키마를 확인해 보았다. 먼저 잠재 수요기업 ‘파이맥스’의 단어는 [표5]와 같고, 값들은 코사인 거리 계수로 잠재 수요기업과 단어 사이의 거리이다.

‘배열검출기형 분광복사계 성능평가 기술’의 잠재 수요기업으로 도출되었던 ‘파이맥스’는 분석단어 ‘광계측’과 설명단어 ‘광계측장비’, ‘LED’, ‘측정’, ‘광측정’, ‘분광’ 등과 공동출현으로 나타났다. 특히 ‘파이맥스’는 ‘배광’, ‘분광분석기’, ‘배광측정기’, ‘고속다채널분광분석기’, ‘3차원배광측정기’ 등과 함께 기술되고 있었다. 이들이 ‘파이맥스’와 의미적 지식 네트워크를 형성하고 있는 스키마이다.

다음으로 ‘초음파를 이용한 진공압력측정센서 및 모듈’의 잠재 수요기업인 ‘프리스케일’의 예고네트워킹분석 결과 얻은 주요단어 및 코사인 거리는 [표6]과 같다.

[표5] 잠재 수요기업 ‘파이맥스’의 스키마

번호	단어	코사인 거리	번호	단어	코사인 거리
1	배광	0.541736	11	정밀광계측	0.14596
2	분광분석기	0.444554	12	Led	0.143569
3	배광측정기	0.442326	13	조명	0.126379
4	고속다채널분광분석기	0.373002	14	광계측	0.125978
5	3차원배광측정기	0.3649	15	광계측시스템	0.104257
6	분광분석	0.321745	16	측정	0.079611
7	엘이디오퍼티컬터치패널기술	0.255377	17	Oled	0.065938
8	광측정	0.233126	18	Lcd	0.058977
9	광계측장비	0.211669	19	분광	0.053838
10	엘이디램프배광측정시스템	0.147442	20	광계측기술	0.042563

[표6] 잠재 수요기업 ‘프리스케일’의 단어

번호	단어	코사인 거리	번호	단어	코사인 거리
1	진공장비	0.948683	8	압력센서	0.344985
2	디지털압력센서	0.9245	9	센서	0.215314
3	디지털센서	0.816497	10	압력측정	0.195069
4	모니터링	0.503953	11	압력	0.188707
5	장비	0.500319	12	측정	0.129555
6	미세전자기계시스템	0.471405	13	진공	0.114291
7	멤스	0.426006			

‘프리스케일’은 ‘진공’, ‘압력’, ‘측정’, ‘센서’의 분석 단어와 ‘압력측정’, ‘압력센서’, ‘장비’의 주요단어의 공동출현 연결을 통해 도출되었다. 잠재 수요기업 ‘프리스케일’과 연관되어 있는 설명단어들을 확인한 결과, ‘진공장비’, ‘디지털압력센서’, ‘디지털센서’, ‘모니터링’, ‘미세전자기계시스템’ 등이 있었다. 설명단어들은 ‘프리스케일’과 의미적 지식 네트워크를 형성하고 있고, 스키마이다.

V. 결론 및 제언

1. 결론

본 연구는 빅데이터 분석을 활용해 기술마케팅의 잠재 수요기업을 발굴하는 방안에 대해 살펴보았다. 이를 위해 문헌연구를 통해 이론적 배경과 연구모형을 제시하고 한국표준과학연구원(KRISS)의 2가지 신기술을 선정하여 데이터를 수집하고 텍스트마이닝과 시맨틱네트워크분석을 실시하였다. 분석 결과로 잠재 수요기업을 발굴하고 홈페이지로 재검증하는 과정을 거쳤다. 잠재 수요기업들의 에고네트워크를 통한 스키마를 추정해 보았다. 빅데이터 분석을 실시한 결과, 세 가지 연구문제가 해결되었다. 첫 번째 연구문제는 연결정도, 연결정도 중심성, 빈도 등을 통해 2개의 기술마다 상위 20개

의 핵심기술을 선정하였다. 두 번째 연구문제는 공동출현 연결망, 연결 강도, 코사인 유사계수 등을 통해 2개 기술의 잠재 수요기업을 발굴하였다. ‘배열검출기형 분광복사계 성능평가 기술’에 대해서는 2개의 분석단어로 수집한 3,225개의 데이터 중에 광학·영상 관련 분야, 분석·측정 장비 분야, 반도체 분야, 광부품업체 등에서 17개의 기업을 발굴하였고, ‘초음파를 이용한 진공압력 측정센서 및 모듈’은 2개의 분석단어에서 1,914개의 데이터를 수집해 제어·계측 분야, 센서·칩 분야, 자동차부품 분야, 공압기기 분야, 반도체 분야에서 총 10개의 기업을 발굴하였다. 세 번째 연구문제는 잠재 수요기업의 에고네트워크분석을 통해 기업명과 의미 있게 연관되어 있는 설명단어들을 확인하였다. 따라서 빅데이터 분석을 실시하여 본 연구가 제기한 3가지 연구문제는 모두 해결되었다.

본 연구의 의의는 기술마케팅의 잠재 수요기업 발굴에 빅데이터를 활용할 수 있음을 실증적으로 보여 주어 빅데이터 분석의 구체적인 활용 분야까지 빅데이터 연구의 범위를 확대하였다는 점이다. 또한 전통적인 마케팅의 목표집단 선정 방법에서 진일보한 기술마케팅의 잠재 수요기업 발굴 방안을 새롭게 제시한 점이다. 지금까지 기술마케팅의 잠재 수요기업 발굴은 불특정 다수기업을 대상으로 기술을 촉진시키는 방법으로 수행되었다. 기술이전 촉진은 특정한 기술의 수요를 조사하여 발굴한 후

기술상담을 실시하는 방식과 기술거래웹사이트 운영, 기술이전 설명회와 같은 판매기술을 공개적으로 소개하는 방식으로 실시되고 있다. 이제 빅데이터 분석을 활용하여 기술마케팅의 잠재 수요기업을 발굴 한 후 그 기업을 대상으로 전략적으로 마케팅 활동을 전개한다면 마케팅의 효율성을 높이면서 기술 이전의 성공 가능성도 높일 수 있을 것이다.

초기단계의 빅데이터 분석 활용 연구라 다음과 같은 한계점들도 있다. 첫 번째, 아직 빅데이터 수집의 기술적 한계로 불필요한 데이터를 많이 수집하고 있다는 점이다. 분석단어의 데이터 수집 과정에서 데이터 수집프로그램이 분석단어를 세분해 인식하는 오류가 있어 세분단어들을 포함하고 있는 데이터들이 많이 수집되어 세분단어들이 분석 결과로 과도하게 계산되어지는 문제가 있다. 두 번째, 포털 사이트의 검색에 의존하는 데이터 수집 프로그램의 문제와 신기술의 특성상 절대적인 데이터의 양이 제한되어 수집한 데이터의 량(volume)이 빅데이터라고 하기에는 부족하다는 점이다. 잠재 수요기업 발굴의 신뢰성을 향상시키기 위해서는 빅데이터의 양적 특성을 충족시킬 필요가 있다. 빅데이터는 데이터의 규모가 커야 하고 텍스트 이외에 이미지, 영상 등 다양한 데이터 형태를 포함하고 있어야 한다. 마지막으로 데이터 수집을 위한 분석단어의 선정이 주관적이고 체계적이지 못하다는 점이다. 분석단어에 따라 수집한 빅데이터가 달라져 분석결과가 완전히 달라질 수 있다. 현재는 기술 설명서와 코더에 의해 선정된 분석단어를 연구자가 기술 개발자와 협의하여 결정하기 때문에 연구의 객관성에 영향을 미칠 수가 있다. 연구의 타당성을 높이기 위해서는 분석단어의 선정을 보다 체계적이고 과학적으로 할 방법이 필요하다.

2. 미래연구를 위한 제언

빅데이터 분석을 기술 마케팅에 활용하기 위해서 몇 가지 제안을 하면 다음과 같다. 첫 번째는 발굴

한 잠재 수요기업을 대상으로 기술 수요조사를 실시하여 기업의 의향을 직접적으로 확인하는 통합적인 마케팅을 실시해 볼 것을 제안한다. 현재 기술 수요조사는 무작위로 이루어지다보니 응답율이 낮고 억지로 떠넘기듯이 조사가 이루어져 신뢰성도 떨어지고 있다. 빅데이터 분석을 통해 잠재 수요기업을 발굴한 다음 이 기업들을 대상으로 기술 수요조사를 실시하면 어느 정도 관심이 있는 상태이기 때문에 보다 적극적인 설문참여를 유도할 수 있고 또한 실제로 잠재 수요기업이 맞지도 확인해 볼 수 있다. 두 번째는 기술개발 이전에 빅데이터 분석을 활용하여 기술개발의 실효율을 높일 것을 제안한다. 기술이전의 효율성을 높이기 위해 기술개발 이후의 빅데이터 분석을 통한 잠재 수요기업 발굴과 함께 기업들의 기술 수요를 사전에 파악하여 기술의 수요예측과 전망을 높이기 위해 기술개발 이전에 빅데이터 분석을 실시할 것이 필요하다. 세 번째는 기술이전의 효과를 더 높이기 위해 잠재 수요기업의 스키마를 활용하여 통합적 마케팅커뮤니케이션을 실시할 것을 제안한다. 소비자들은 자신의 스키마에 의존해 정보를 수용하고 기억을 활성화하고 있기에 잠재 수요기업과 연관성이 있는 단어들을 사용해 커뮤니케이션을 실시하면 효과를 극대화할 수 있을 것이다. 네 번째는 빅데이터의 가치를 훼손시키지 않고 정제화의 시간도 단축시킬 수 있도록 텍스트마이닝 프로그램의 개발을 제안한다. 현재는 자연어처리와 형태소분석의 정확도가 낮아 거의 수작업에 의존하고 있기에 시간과 인력이 너무 많이 투입되고 있다. 신뢰성 높고 빠른 분석을 위해 한글 자연어처리와 형태소분석에 최적화된 프로그램의 개발이 필요하다.

참 고 문 헌

권선필(2008). “컴퓨터 활용 질적자료 분석을 통한 행정-정치-경영 담론 분석.” *한국행정학보*, 42(1), 345-364.

김경용(1994). *기호학이란 무엇인가?*. 서울, 민음사

김광철, 장병원(2004), *영화사전*, 서울, MEDIA2.0

김상락, 강만모, 박상무(2012), “빅데이터가 여는 미래의 세상,” *정보과학회지*, 30(6), 18-24.

김승윤(2012), “Big Data 최근 글로벌 동향과 이슈-국가 경쟁력 확보 위한 공공부문에서의 Big Data 활용사례”, *Issue & Trend*, KT 경제경영연구소

김정숙(2012). “빅데이터 활용과 관련기술 고찰.” *한국콘텐츠학회지*, 10(1), 34-40.

박영원(2003), *광고디자인 기호학*, 범우사

박한우(2012), “소셜 여론조사의 실제와 과제-‘저비용 고효율’ SNS로 여론 읽기,” 월간<신문과 방송> 7, 84-88.

박한우, 루츠 레이테스도르프(2004). “한국어의 내용분석을 위한 KrKwic 프로그램의 이해와 적용: Daum.net에서 제공된 지역혁신에 관한 뉴스를 대상으로.” *Journal of The Korean Data Analysis Society*, 6(5), 1377-1388.

서일원, 전채남, 이덕희(2013), “시맨틱네트워크분석을 이용한 원천기술 분야의 잠재적 기술 수요 발굴 기법에 관한 연구,” *기술혁신연구*, 21(1), 279-301.

손동원(2002). *사회 네트워크 분석*. 경문사

심홍진, 김용찬, 손해영, 임지영(2011). “언어 네트워크 분석을 통한 스마트폰과 소셜미디어 이용자의 미디어 이용행태에 관한 탐색적 연구 : 성별과 연령에 따른 이용행태 인식과 평가를 중심으로.” *한국방송학보*, 25(4), 82-138.

안창원, 황승구(2012). “빅데이터 기술과 주요 이

슈.” *정보과학회지*, 30(6), 10-17.

오정근(2010). *기호학적 접근방법에 의한 축제의 의미와 의미구조*. 서울, 한국학술정보.

유민호(2011), “데이터를 돈으로 바꿔라.” *주간조선*, 2144, [http://weekly.chosun.com/client/news/viw.asp?nNewsNumb=002144100007&ctcd=C06].

이만재(2011). “빅데이터와 공공데이터 활용”, *Internet and Information Security*, 2(2), 47-64.

정근하(2011), *텍스트마이닝과 네트워크분석을 활용한 미래예측 방법 연구*, 한국과학기술기획평가원

정성훈(2012). “이슈프레임분석을 위한 전통적 내용분석과 언어네트워크분석 방법 비교”, 한양대학교 대학원 석사학위 논문

함유근, 채승병(2012). *빅데이터, 경영을 바꾼다*. 서울, 삼성경제연구소

한기하, *기호학*, [http://www.cyworld.com/Agari29/4306483].

Barabasi, A. L., Albert, R. (1999), “Emergence of Scaling in Random Networks,” *Science*, 286, 509 - 512.

Beyer, M. A., Laney, D. (2012). “The Importance of ‘Big Data’: A Definition”. Gartner, [http://www.gartner.com/id=2057415].

Borgatti, S. P., Everett, M. G. & Freeman, L. C. (2002), *Ucinet for Windows: Software for Social Network Analysis*, Massachusetts, Analytic Technologies

Danowski, J. (1993), *Network analysis of message contents*, In G. Barnett & W. Richards(Eds.), *Progress in communication sciences*, 12, 197-222. Norwood, NJ: Ablex.

Doerfel, M. L., Barnett, G. A. (1999), “A

- semantic network analysis of the International Communication Association," *Human Communication Research*, 25(4), 589 - 603.
- Freeman L. C., Borgatti, S. P., & White, D. R. (1991), "Centrality in valued graphs: A measure of betweenness based on network flow." *Social Networks*, 13, 141-154.
- Hanneman, R. A., & Riddle, M. (2005), *Introduction to socialnetwork methods*, Retrieved February 20, 2006, [http://faculty.ucr.edu/~hanneman].
- Hinchcliffe, D. (2011), *Social Business Intelligence: Positioning a Strategic Lens on Opportunity*, [http://dachisgroup.com/2011/08/social-business-intelligence-positioning-a-strategic-lens-on-opportunity].
- IDC (2012), "Worldwide Big Data Technology and Services 2012-2015 Forecast," *Informatica*, Big Data Unleashed, [http://www.informatica.com/downloads/1601_big_data_wp.pdf]
- Jonassen, D. H., Beissner, K., & Yacci, M. (1993). *Structural Knowledge*. New Jersey, Hillsdale.
- Khalifa, M. & Liu, V. (2006). "Semantic network discussion representation: Applicability & some potential benefits," *Professional Communication*, 49(1), 69~81.
- Park, H. W., & Leydesdorff, L. (2013). "Decomposing Social and Semantic Networks in Emerging"BigData" Research," *Journal of Informetrics*.
- Russom, P. (2011), *Big Data Analytic Report : Big Data Analytics*, TDWI.
- Wang, W. & Rada, R. (1998), "Structured hypertext with domain semantics," *ACM Trans. Inform. Syst.*, 16, 372~412.
- Wasserman, S., Faust, K. (1994), *Social Network Analysis: methods and Applications*, New York, Cambridge University Press.
- Wikipedia, *Big data*, [http://en.wikipedia.org/wiki/Big_data]

Analyzing the Bigdata for Practical Using into Technology Marketing : Focusing on the Potential Buyer Extraction

ChaeNum Jun · Il Won Seo

Abstract

Even though the study of the Bigdata is getting advanced, the topics are limited fundamental concept, trend, technical situation or the practical possibility of the Bigdata. The purposes of this research are to expand the study of the Bigdata and to use it in technology marketing practically. Also we found a few buyers which have potential possibility of technology marketing by using Bigdata. One reasons that we choose Bigdata for technology marketing is the necessity of marketing in the technical business area is getting higher. The other reason is that finding potential buyers is getting more important because of the nature of technical business. For these, we set-up the research problems like “What are the verified main techniques through the Bigdata analyzing?, Where are the extracted potential buyers through the Bigdata analyzing?” and “What are the keywords which are connected to the potential buyers?”

We choose two at the new techniques of Korea Research Institution of Standards and Science(KRISS) and three coders discussed to choose four analyzing words. We collected the Bigdata on the internet, and executed textmining and semantic network analysis with it. After that, we reexamined them by confirming the schema of the potential buyer and the homepage. First, we checked 20 main techniques per the two techniques by degree, degree centrality, and frequency. Second, we found the answer of this study by finding the potential buyers which has the two techniques through co-occurrence frequency, network intensity and cosine similarity coefficient. Extract 17 buyers at ‘Performance characterization of array-type spectroradiometers’ and 10 buyers were extracted at ‘Pressure measuring system for vacuum chamber using ultrasonic wave’. Finally, we need to check the keywords related to the corporate name through analyzing ego-network of

ChaeNum Jun THE IMC
Il Won Seo KRISS

potential buyers. Therefore the results are proved completely through analyzing the Bigdata.

This study is full of significance in the way that conceptual study of analyzing Bigdata goes a step further to the application of practical use to find potential buyers. This research contributes to the Bigdata study by expanding the range of the study through offering practical way of using in technology marketing area.

Key Words : Bigdata, Bigdata Analysis, Semantic Network Analysis, Technology Marketing, Potential Buyer