

특허분석을 통한 빅데이터 분석 플랫폼 기술 개발 동향

노승민
성결대학교 멀티미디어공학부

Big Data Analysis Platform Technology R&D Trend through Patent Analysis

Seungmin Rho

Department of Multimedia, Sungkyul University, Anyang-si, Korea

요약 모바일 SNS, 스마트 기기의 급성장과 ICT(통신기술) 패러다임의 변화는 라이프 스타일의 변화와 함께 데이터가 폭발적으로 증가하는 결과를 초래하였다. 한편, 대용량의 데이터의 사용은 비용이나 기술적 측면에서 많은 어려움을 초래하였다. 그러나, 빅데이터는 하둡과 같은 효율적인 빅데이터 처리 플랫폼 기술의 등장으로 주목을 받기 시작하였다. 본 논문에서는 주요 시장 국가의 특허 분석을 통해 빅데이터 플랫폼 연구 및 개발 동향을 살펴보고자 한다. 특히, 2010년 12월을 기준으로 4개국에 출원 및 등록되어 있는 2,568건을 대상으로 분석을 진행하였다.

주제어 : 빅 데이터, 특허, 동향분석, 하둡, 맵 리듀스, 플랫폼

Abstract The ICT (information and communication technology) paradigm shift, including the burgeoning use of mobile, SNS, and smart devices, has resulted in an explosion of data along with lifestyle changes. We have thus arrived at the age of big data. In the meantime, a number of difficulties have arisen in terms of cost or on the technical side with respect to the use of large quantities of data. However, big data has begun to receive attention with the advent of efficient big data technologies such as Hadoop.

In this paper, we discuss the patent analysis of big data platform technology research and development in major countries. Especially, we analyzed 2,568 patent applications and registered patents in four countries on December 2010.

Key Words : Big data, patent, R&D trend, Hadoop, Mapreduce, platform

1. 서론

인터넷, 웹, 모바일, 스마트 기기, 센서 등 정보화의 발전과 ICT (information and communication technology) 패러다임의 변화는 인류의 생활에 큰 변화를 일으키고,

더불어 데이터의 폭증을 가져오게 되었다. 이로 인해 우리는 방대하고 다양한 데이터들이 실시간으로 쏟아지는 데이터 시대에 살게 된 것이다. International Data Corporation은 2011년 전 세계 디지털 정보량은 1.8 제타바이트에 달하는 ‘제타바이트 시대’에 진입하였고, 전 세

Received 11 July 2014, Revised 28 August 2014

Accepted 20 September 2014

Corresponding Author: Seungmin Rho

(Department of Multimedia, Sungkyul University, Anyang-si, Korea)

Email: smrho@sungkyul.edu

ISSN: 1738-1916

© The Society of Digital Policy & Management. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

계의 데이터양은 2년마다 두 배씩 증가하는 것으로 제시하였다. 그동안 대량의 데이터들(빅데이터)을 처리 및 활용하기 위해서는 기술적으로나 비용적인 측면에서 많은 어려움이 있었다. 그러나 최근 빅데이터를 효율적으로 분석, 처리해낼 수 있는 기술들이 등장하면서 실시간으로 빅데이터를 분석하고 관련 데이터를 융복합하여 가치 있는 의미를 찾아낼 수 있게 되면서 빅데이터가 중요한 디지털 자원으로 주목받기 시작하였다[1][7].

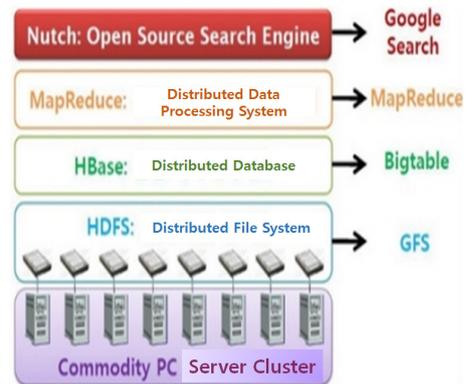
빅데이터란 좁은 의미로는 “방대하고 복잡도가 매우 높아 기존의 데이터베이스 도구로는 처리 및 관리가 불가능한 데이터 세트[1]”로 정의된다. Beyer와 Laney [2]는 좀더 넓은의미에서, 빅데이터를 “큰 볼륨, 빠른 속도 및 높은 다양성(3Vs; volume, velocity, variety)의 특성을 가지고 있어 관리가 어려운 데이터 및 이들 데이터의 축적·처리·분석을 위한 기술과 나아가서는 데이터를 분석해 향상된 의사결정과 유용한 의미 및 통찰, 프로세스 최적화를 가능하게 하는 인재와 조직을 포함하는 포괄적인 개념”으로 정의한다[3].

본 논문에서는 클라우드 기반 빅데이터 분석 협업 플랫폼 기술을 개발함에 있어, 다양한 위치의 데이터 접근 및 수집이 가능하도록 하는 데이터 가상화, 정형/비정형 대용량 데이터의 저장 및 검색에 관한 대용량 데이터 저장관리, 분산 저장된 데이터 대상 병렬 연산 처리가 가능하도록 하는 분산 병렬처리를 포함하는 빅데이터 인프라 기술과, 분석 알고리즘의 병렬 수행이 가능하도록 하는 병렬 분석 엔진, 예측 분석, 평판 분석 등 다양한 데이터 융합이 가능하도록 하는 고급 분석 기술, 수학적 모델링 기반의 데이터 분석이 가능하도록 하는 자율 데이터 분석, 분석 결과의 신뢰성을 검증하는 신뢰성 검증 기술을 포함하는 빅데이터 분석엔진 기술과, 다중사용자가 공유하는 개방형 빅데이터 서비스 환경을 제공하는 빅데이터 마켓플레이스, 빅데이터를 활용한 서비스 개발 지원 환경을 제공하는 분석 서비스 모델링, 빅데이터 분석 결과를 사용자 요구사항에 맞춤화하여 가시화하는 분석결과 가시화를 포함하는 빅데이터 공유 환경 기술에 대하여 특허분석을 통해 빅데이터 플랫폼 기술의 발전 및 연구 방향을 전망한다.

2. 빅데이터 처리 및 분석 기술

2.1 하둡

하둡(Hadoop)[3][5]은 오픈소스 분산처리기술 프로젝트로, 현재 정형/비정형 빅 데이터 분석에 가장 선호되는 솔루션이라고 할 수 있다. 실제로 야후와 페이스북 등에 사용되고 있으며, 채택하는 회사가 늘어나고 있다. 주요 구성요소로 [Fig. 1]의 하둡 분산 파일 시스템인 HDFS(Hadoop Distributed File System), Hbase, MapReduce가 포함된다. HDFS와 Hbase는 각각 구글의 파일 시스템인 GFS(Google File System)와 빅 테이블(Big Table)의 영향을 받았다. 기본적으로 비용효율적인 x86 서버로 가상화된 대형 스토리지(HDFS)를 구성하고, HDFS에 저장된 거대한 데이터셋을 간편하게 분산처리할 수 있는 Java 기반의 MapReduce 프레임워크를 제공한다. 이외의 Hadoop을 기반으로 한 다양한 오픈소스 분산처리 프로젝트가 존재한다.



[Fig. 1] Hadoop structure and distributed technology of Google

2.2 MapReduce

맵리듀스(MapReduce)[4][6]는 대용량 데이터 분석에 대한 분산 및 병렬 처리를 지원하기 위해 개발된 프레임워크이며, 프로그래머가 분산 처리에 대한 지식이 없더라도, 맵(map)과 리듀스(reduce)라는 함수만 구현하면 클러스터를 활용할 수 있도록 돕는다.

맵리듀스는 최근 쉽게 접할 수 있는데, 하둡(Hadoop)과 같이 오픈 소스로 구현된 경우도 있고, Greenplum 이

나 Aster Data와 같은 MPP (Massively Parallel Processing) 기반의 대용량 데이터베이스 시스템에서도 맵리듀스 프레임워크를 구현하여 제공하고 있다. 맵리듀스를 이용한 데이터 분석도 많이 시도되고 있는데, 최근에는 맵과 리듀스 함수를 직접 구현하지 않고, 선언적인 언어를 이용하여 데이터 분석을 지원하는 도구들도 제공되고 있다[5].

3. 빅데이터 분석 플랫폼의 특허동향

3.1 기술분류체계 및 기준

본 논문에서는 클라우드 기반 빅데이터 분석 협업 플랫폼 개발 기술의 특허 분석을 위해 빅데이터 인프라, 빅데이터 분석엔진, 빅데이터 공유 환경 기술로 중분류하였고, 빅데이터 인프라 기술을 데이터 가상화, 대용량 데이터 저장관리, 분산 병렬처리, 빅데이터 분석엔진 기술을 병렬 분석 엔진, 고급 분석, 자율 데이터 분석, 신뢰성 검증, 빅데이터 공유 환경 기술을 빅데이터 마켓플레이스, 분석 서비스 모델링, 분석결과 가시화 기술로 각각 소분류 하였으며, 심층분석(정성분석)시의 기술 분야를 동

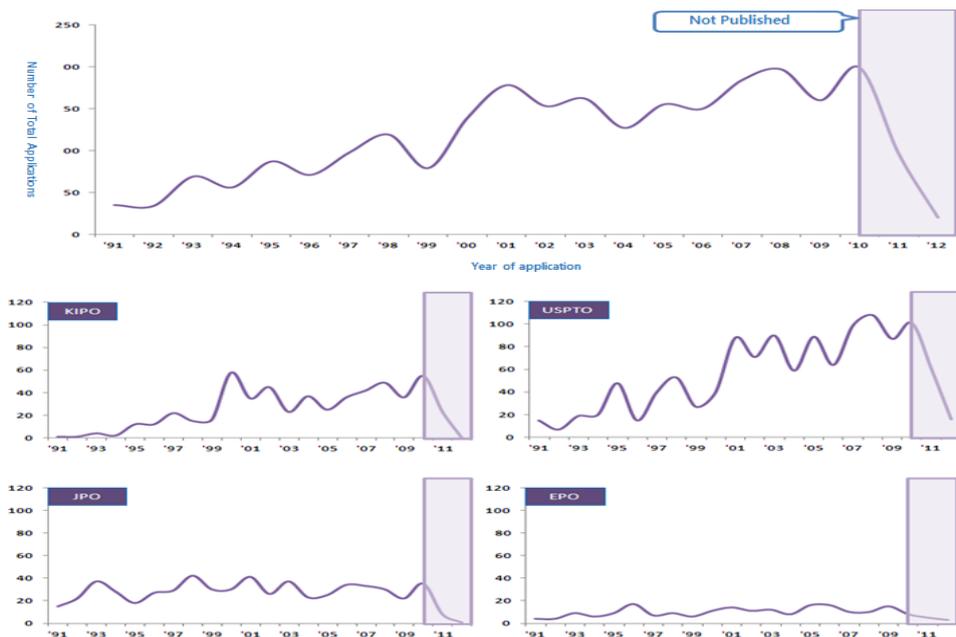
일하게 적용하였다.

3.2 국가별 특허동향

클라우드 기반 빅데이터 분석 협업 플랫폼 개발 기술 분야의 연도별 전체 특허동향을 살펴보면, 거시적인 관점에서 특허출원건수는 1991년부터 최근까지 증가와 감소가 반복되는 추세를 보이고 있으며, 전체적으로 증가하는 추세를 보이고 있다. 미공개 특허가 존재하는 2010년 및 2011년 구간도 빅데이터 비즈니스 활용 요구 증대되면서 빅데이터 처리 및 분석 기술 분야가 최근 관심 기술임을 고려해 볼 때, 증가추세 곡선이 될 것으로 예측된다.

[Fig. 2] 에서 한국의 경우에는 1991년부터 완만한 증가세를 보이다가 2000년에 급증하는 양상을 보이고 있으며, 이때 58건의 특허출원이 이루어지면서 가장 많은 건수를 보이고 있고, 이 이후로 등락을 거듭하면서 다소 감소하는 추세를 보이다가, 최근 다시 증가하는 양상을 보이고 있다.

미국은 다른 국가들에 비해 상당히 많은 수의 특허가 출원되었고, 1991년부터 특허출원건수의 증가와 감소가 반복되다가 2001년에 88건으로 특허출원건수가 급증하



[Fig. 2] Status of annual patent application

였으며, 이 이후로도 최근까지 특허출원건수의 증가와 감소가 반복되는 양상을 보임. 2008년에는 108건의 특허출원으로 가장 많은 건수를 나타내고 있으며, 대체적으로 증가하는 추세를 보인다.

일본은 1991년부터 최근까지 50건 미만의 특허출원건수에서 등락을 거듭하고 있으나, 특허 출원이 꾸준히 이루어지는 편이고, 2001년에 41건으로 가장 많은 특허출원을 하는 것으로 나타났다.

유럽에서의 매년 특허출원건수는 20건 미만으로 다른 국가들에 비해 미진한 편이며, 1996년에 17건의 특허출원이 이루어지면서 가장 많은 특허출원건수를 보이고 있고, 최근 감소하는 추세를 보인다.

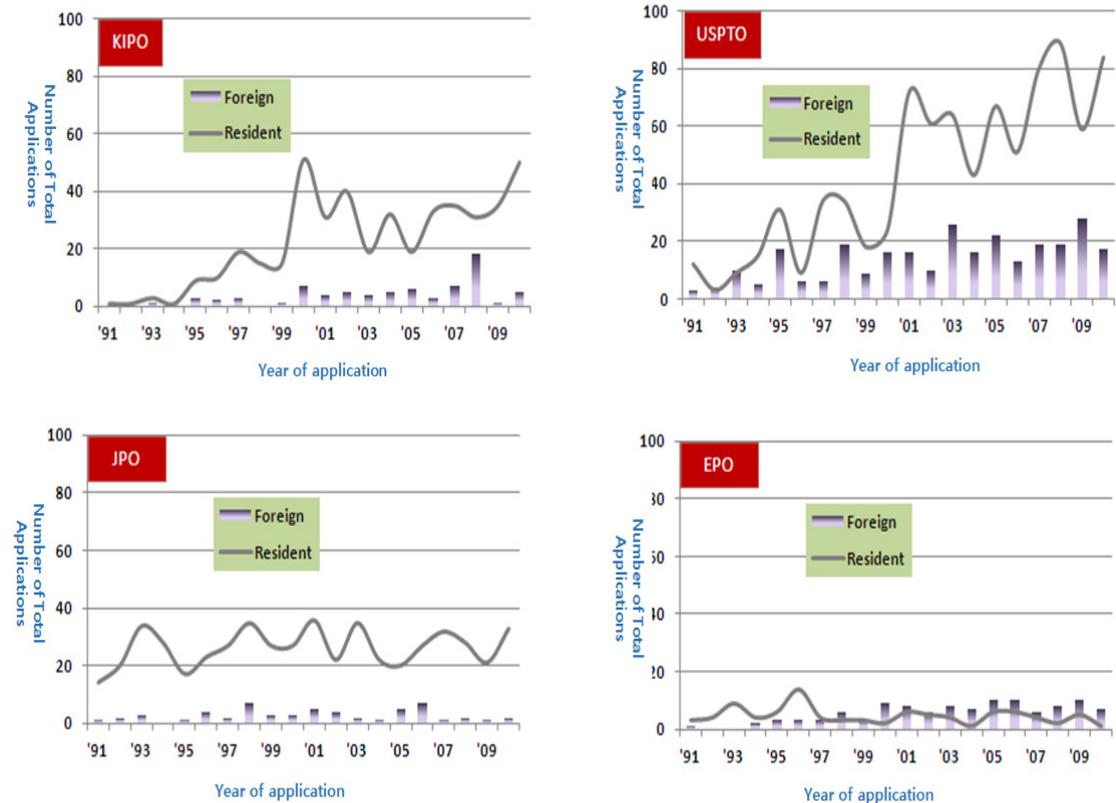
3.3 주요국의 특허출원 현황

클라우드 기반 빅데이터 분석 협업 플랫폼 개발 기술 분야의 국가별/출원인 국적별 특허동향을 살펴보면, 미

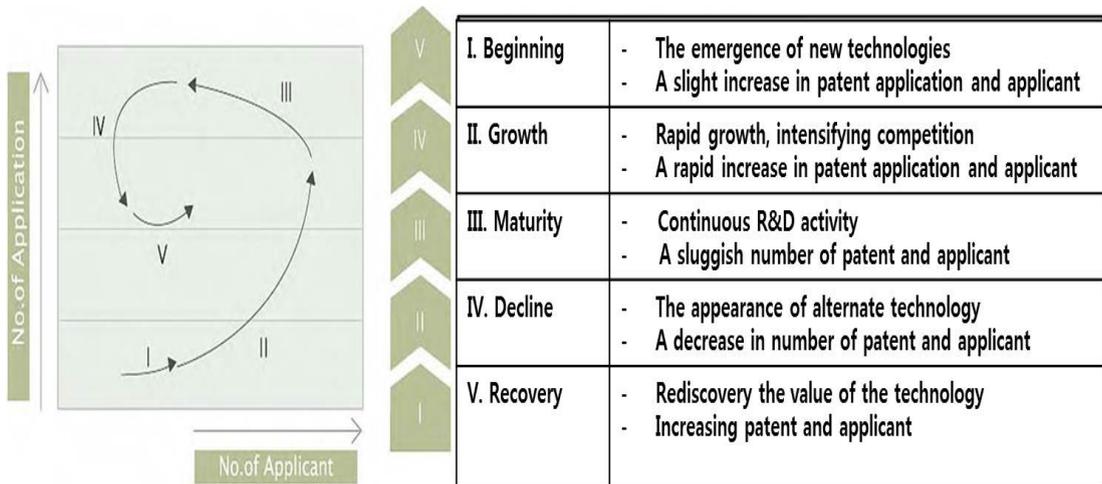
국에서의 출원이 전체 분석대상 국가 출원규모의 절반 가까이 차지하는 것으로 나타나, 이 기술분야의 연구개발은 대부분 미국에서 주도되고 있는 것으로 보인다.

[Fig. 3]의 주요국의 내·외국인 특허출원현황을 살펴보면, 유럽에서의 외국인 점유율이 54%로 가장 높는데, 이 외국인 출원건수 중 51%를 미국인 국적의 출원인이 출원하고 있는 것으로 나타나, 역시 미국이 연구개발을 주도하고 있음이 확인된다.

한국과 일본에서의 외국인 점유율이 14, 20%로 나타나 외국인보다는 내국인에 의한 특허활동이 활발한 것으로 나타났으며, 한국에서는 미국인이 외국인 중 46%를, 일본에서는 미국인이 외국인 중 78%를 차지하고 있는 것으로 나타나, 미국에서 활발한 연구개발결과를 세계 각국으로 출원하여 권리화하고 있는 것으로 판단된다. 또한, 미국에서는 내국인이 76%를, 외국인이 24%를 차지하고 있는데, 이 중 일본 국적이 60%를 차지하고 있어,



[Fig. 3] Patent application status of major domestic and foreign nation by year



[Fig. 4] Criteria of the growth stage technology market

미국 다음으로 일본에서 활발한 연구개발 활동이 이루어지고 있는 것으로 판단된다. 이러한 원인은 인터넷이 미국에서부터 시작되었으며, 인터넷에서의 대규모 데이터를 수집하여 처리, 가공, 관리하도록 하는 기술 또한, Google, Microsoft 등과 같은 미국의 대규모 인터넷 업체들이 주도하고 있기 때문으로 풀이된다.

한국에서 내국인의 특허출원건수는 2000년까지 증가하는 양상을 보이다가, 이 이후의 분석구간에서 미약한 감소세를 보이며, 외국인에 의한 특허출원건수는 2000년대에 증가와 감소가 반복되고 있다.

미국에서 내국인의 특허출원건수는 분석구간 전체에 걸쳐서 대체적으로 증가하는 양상을 보이며, 외국인에 의한 특허출원건수는 완만한 증가세를 보이고 있다.

일본에서 내국인과 외국인의 특허출원건수는 분석구간 전체에 걸쳐서 완만하게 유지되는 양상을 보인다.

유럽에서 내국인에 의한 특허출원건수는 1990년대 말 이후로 감소하고, 외국인의 특허출원건수는 2000년부터 서서히 증가하는 양상을 보인다.

4. 빅데이터 기술의 세부 특허 분석

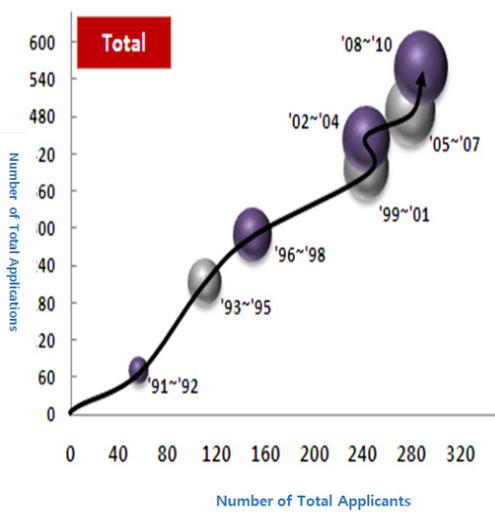
기술시장 성장단계 그래프는 총 5가지 해석기준으로 구분할 수 있는데 출원의 건수에 따라서 태동, 성장, 성숙, 쇠퇴, 회복으로 기술시장을 분석할 수 있다. [Fig. 4]

는 기술시장 성장 단계의 해석 기준을 나타낸다.

[Fig. 5] 는 클라우드 기반 빅데이터 분석 협업 플랫폼 개발 기술 분야의 해당 국가의 기술 위치를 포트폴리오로 나타낸 그래프이다. 전체 출원 중 최근의 출원 동향을 7개의 구간으로 나누어 각각의 구간별 특허 출원인 수 및 출원 건수를 나타내어 특허 출원 동향을 통한 기술의 위치를 살펴볼 수 있다. 본 논문에서는 각 구간을 1구간(1991년VI~1992년), 2구간(1993년~1995년), 3구간(1996년~1998년), 4구간(1999년~2001년), 5구간(2002년~2004년), 6구간(2005년~2007년), 7구간(2008년~2010년)으로 나누었다.

[Fig. 6] 에서는 전 세계 기술 위치를 포트폴리오로 나타낸 것으로 클라우드 기반 빅데이터 분석 협업 플랫폼 개발 기술 분야는 1구간(1991년~1992년)에서 태동기를 나타내고 있으며, 2구간(1993년~1995년) 내지 4구간(1999년~2001년)에서 성장기 단계를 나타내고 있음. 5구간(2002년~2004년)에서는 성숙기 단계를 보이고 있으며, 6구간(2005년~2007년)과 7구간(2008년~2010년)에서 출원인 수와 출원 건수가 증가하고 있고, 이 기술 관련 분야의 전체 및 해당 국가의 기술 위치를 포트폴리오로 나타낸 것으로 전체 출원 중 최근의 출원 동향을 7개의 구간으로 나누어 각각의 구간별 특허 출원인 수 및 출원 건수를 나타내어 특허 출원 동향을 통한 기술의 위치를 살펴볼 수 있다.

1구간(1991년~1992년)에서 출원 건수와 출원인의 수



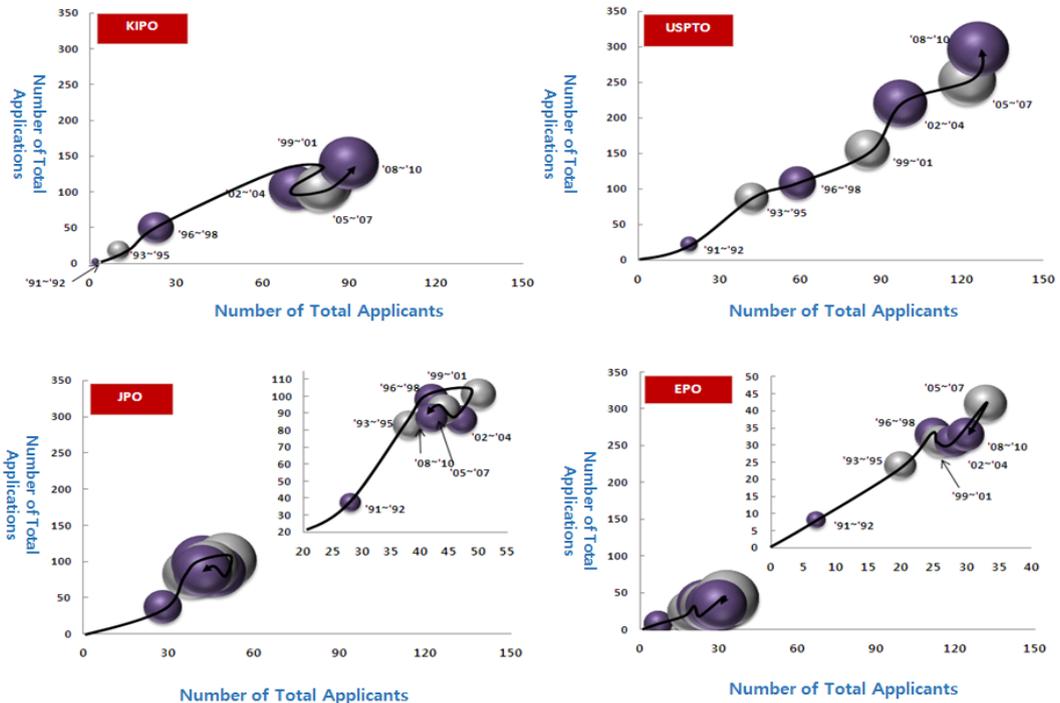
[Fig. 5] Growth stage of technology market

가 완만하게 증가하는 태동기 단계에 있으며, 2구간(1993년~1995년) 내지 4구간(1999년~2001년)까지 출원 건수 및 출원인 수가 크게 증가하고 있어 이 시기에 급격한 기

술 개발이 이루어진 것으로 분석됨. 이후 5구간(2002년~2004년)에서 출원인의 수는 약간 감소한 반면 출원 건수는 다소 증가하고 있어 성숙기 단계에 있고, 이후 6구간(2005년~2007년)과 7구간(2008년~2010년)에서 출원 건수와 출원인의 수가 증가하고 있어 다시 기술 개발이 이루어진 것으로 분석된다.

포트폴리오로 나타난 한국특허(KIPO)의 기술위치는 1구간(1991년~1992년) 내지 3구간(1996년~1998년)까지 출원 건수와 출원인의 수가 완만하게 증가하는 태동기를 거쳐 4구간(1999년~2001년)에서 출원 건수 및 출원인 수가 크게 증가하는 성장기 단계에 있는 것으로 보이며, 5구간(2002년~2004년)에서 출원 건수 및 출원인 수가 감소하고 있어 쇠퇴기 단계에 있는 것으로 보이고, 6구간(2005년~2007년) 및 7구간(2008년~2010년)에서는 출원 건수 및 출원인 수가 증가하고 있다.

미국특허(USPTO)의 기술위치는 1구간(1991년~1992년)에서 태동기에 있다가 2구간(1993년~1995년)부터 성장기 단계에 접어들어 그 경향이 현재까지 지속적으로 유지되고 있는 것으로 보인다.



[Fig. 6] Growth stage of technology market by nation

일본특허(JPO)의 기술위치는 1구간(1991년~1992년)에서 태동기에 있다가 2구간(1993년~1995년) 내지 4구간(1999년~2001년)까지 성장기에 접어들었으며, 5구간(2002년~2004년)에서 출원건수와 출원인의 수가 감소하는 양상을 보이고, 유럽특허(EPO)의 기술위치는 2구간(1993년~1995년) 내지 3구간(1996년~1998년)까지 출원건수와 출원인의 수가 완만하게 증가하는 성장기에 있고, 4구간(1999년~2001년)에서 출원 건수가 다소 감소하는 양상을 보이고 있으며, 5구간(2002년~2004년)과 6구간(2005년~2007년)에서 출원 건수와 출원인의 수가 다소 증가하다가 7구간(2008년~2010년)에서 출원건수 및 출원인 수가 감소하는 쇠퇴기 단계에 있는 것으로 보임. 또한, 해당 구간의 출원 건수와 출원인 수가 급격히 감소하는 것으로 보아 해당 구간의 기술개발 빈도가 크게 줄어든 것으로 분석된다.

5. 결론

본 논문에서는 빅데이터 플랫폼 기술의 특허 분석을 통하여 각각의 플랫폼 기술의 동향을 살펴보았다. 분석 데이터에 따르면 한국, 미국, 유럽의 경우에는 최근에 출원이 다시 증가하는 추세이고, 일본은 반대로 최근에 출원이 감소하는 추세를 보이고 있다. 그러나, 출원인 별로 살펴보면 3국 패밀리수(미국·일본·유럽 공동 출원 특허수), 최근 3년간의 특허출원 증가율을 비교분석한 결과, 주요출원인들은 전반적으로 주요 시장국 모두에서 활발한 특허활동을 하고 있는 것으로 나타났다. 세부 기술별로 살펴보면 한국은 빅데이터의 저장관리, 분석기술에서 중점적으로 출원이 이루어지고 있는 반면 미국은 빅데이터의 모든 분야에서 골고루 출원이 이루어지고 있어 빅데이터 플랫폼 기술을 주도하고 있는 것으로 나타났다. 이는, IBM, YAHOO, ORACLE MS등과 같은 세계적인 IT 업체가 빅데이터의 다양한 기술에 집중적으로 투자를 하고 있기 때문이다. 이러한 추세에 발맞추어 한국의 산학연에서는 대량의 데이터를 수집하고 이를 심층적으로 분석함으로써 새로운 비즈니스 가치를 발굴하는 데이터 사이언티스트의 양성이 조속히 이루어져야 할 필요가 있으며, 정부는 절대적으로 부족한 교육인프라 확충, 실무 능력 배양을 위한 지원강화를 통해 데이터 사이언티스트 양성을 위한 정책적 지원을 확대해야 할 것이다.

REFERENCES

- [1] White T. Hadoop: the definitive guide. Beijing: O'Reilly; 2012. 2. Beyer MA, Laney D. The importance of 'big data': a definition. Stamford: Gartner; 2012.
- [2] Beyer MA, Laney D. The importance of 'big data': a definition. Stamford: Gartner; 2012 [cited 2014 Apr 15]. Available from: <https://www.gartner.com/doc/2057415/importance-big-data-definition>.
- [3] McKinsey, "Big Data: The Next Frontier for Innovation, Competition, and Productivity," McKinsey & Company, 2011.
- [4] J. Dean and S.Ghemawat, "MapReduce:Simplified Data Processing on Large Clusters," Sixth Symposium on Operating System Design and Implementation, pp.137-150, Dec.2004.
- [5] K.Shvachko and H.Kuang, "The Hadoop distributed FileSystem," Yahoo!, 2010.
- [6] J. Kelly et al., "Bigdata Vendor Revenue and Market Forecast 2012-2017," Wikibon, Feb. 19th. 2013.
- [7] Sitaram Asur, Bernardo A. Huberman, "Predicting the Future With Social Media," Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference, vol.1, pp.492-499, 2010.

노 승 민(Rho, Seungmin)



- 2008년 8월 : 아주대학교 정보통신 공학과(공학박사)
- 2008년 10월 ~ 2009년 9월 : Carnegie Mellon University (박사 후연구원)
- 2009년 10월 ~ 2012년 2월 : 고려대학교 전자공학부 연구교수
- 2012년 3월 ~ 2013년 2월 : 백석대학교 정보통신학부 조교수
- 2013년 3월 ~ 현재 : 성결대학교 멀티미디어공학부 조교수
- 관심분야 : 멀티미디어 검색/추천 및 응용, 빅데이터 분석
- E-Mail : smrho@sungkyul.edu