

DATA INDUSTRY

WHITE PAPER

2 0 1 6

2016 데이터산업 백서



DATA INDUSTRY WHITE PAPER 2016

부문별 데이터산업 동향

제1장 데이터 서비스 동향

- 제1절 공공데이터
- 제2절 기상정보
- 제3절 보건의료정보
- 제4절 신용정보
- 제5절 특허정보
- 제6절 교육·훈련정보

제2장 데이터 솔루션 동향

- 제1절 데이터 수집
- 제2절 데이터 설계
- 제3절 DBMS
- 제4절 데이터 관리
- 제5절 데이터 보안
- 제6절 데이터 품질
- 제7절 데이터 분석

제3장 데이터 컨설팅 및 구축 동향

- 제1절 데이터 설계 컨설팅
- 제2절 데이터 품질 컨설팅
- 제3절 데이터 거버넌스 컨설팅
- 제4절 데이터 이행 컨설팅
- 제5절 데이터 분석 컨설팅
- 제6절 비즈니스 인텔리전스 구축
- 제7절 오픈 데이터 구축

제4장 데이터 산업기술 동향

- 제1절 지능정보 기술
- 제2절 빅데이터 기술
- 제3절 클라우드 기술
- 제4절 사물인터넷 기술



제1장 데이터 서비스 동향

제1절 공공데이터

1. 공공데이터 개요

정보 사회는 정보의 생성, 유통, 사용, 통합 및 조작이 경제적, 정치적, 문화적 활동의 핵심으로, 디지털 정보와 이를 매개로한 소통 기술이 정치, 경제, 교육 등 사회 전반을 변화시키는 원동력이 되는 사회를 말한다.³⁵⁾ 단적으로 이전 산업 사회와 비교했을 때 정보 사회는 경제 활동의 중심이 지식, 정보에 있다.

최근에는 ‘데이터’가 새로운 경제, 정보 사회를 대표하는 핵심 요소로 부각되고 있다. 링크드 데이터, 개방형 데이터, 빅데이터, 정부 2.0(Gov2.0) 등이 새로운 데이터 사회를 대표하는 새로운 용어들로 이 용어는 공통적으로 가공된 정보가 아닌 이전 단계의 데이터 활용에 대한 가능성에 중점을 두고 있다. 정보화 사회에서 데이터는 흔히 ‘새로운 오일(data is the new oil)’로 지칭하며, 경제적 파급 효과의 분석을 위한 노력이 계속되고 있다. 데이터 경제는 본격적으로 데이터 주도의 새로운 사회 질서를 나타내며, ‘데이터’가 가져올 수 있는 실질적인 혜택과 구현 등에 관한 논의를 중심으로 확산되고 있다.

공공데이터는 이런 새로운 데이터 중심 정보 생태계 초기에 양적, 질적 인프라 구축의 가장 핵심적인 영역을 차지하고 있으며 공공데이터 중에서도 정부 데이터가 가장 큰 역할을 담당하고 있다.

³⁵⁾ Hilbert, M. (2015). Digital Technology and Social Change [Open Online Course at the University of California]

2. 개방형 정부 데이터

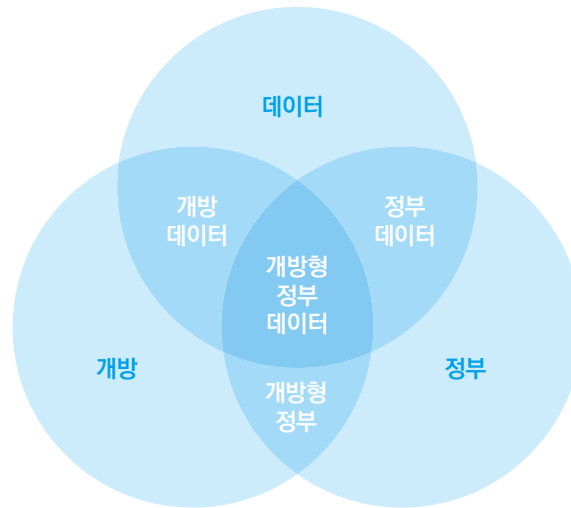
‘공공데이터’란 데이터베이스, 전자화된 파일 등 공공기관이 법령 등에서 정하는 목적을 위하여 생성 또는 취득하여 관리하는 광(光) 또는 전자적 방식으로 처리된 자료 또는 정보³⁶⁾를 말한다.

개방형 데이터는 누구나 자유롭게 활용할 수 있는 데이터로 데이터의 사용, 재발행에 있어서 법률적·기술적 제한이 없는 것³⁷⁾을 말한다.

공공데이터의 주체는 공공기관이다. 특히, 정부가 구축한 정부 데이터가 절대 다수를 차지하고 있다. 열린 데이터는 주요 선진국의 열린 정부 선언(Open Gov.) 및 정책 추진과 맞물려 세계적인 확산 추세를 맞게 되었고, 대표적인 것이 미국과 영국 정부의 열린 정부 선언이다.

공공 데이터, 개방 데이터, 정부 데이터는 새로운 경제적 파급 효과를 만들어 낼 수 있는 데이터를 핵심으로 하는 용어로 개방형 정부 데이터(Open Government Data) 등 합성어로 범용적으로 사용되고 있다.

그림 3-1-1 개방형 정부 데이터 관계 개념도



※ 출처: <https://scasia.co/2015/09/01/governments-data-will-be-made-open-for-all-to-ensure-transparency/>

³⁶⁾ 공공데이터의 제공 및 이용 활성화에 관한 법률, [시행 2016.4.7.] [법률 제13723호, 2016.1.6. 일부개정]

³⁷⁾ Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., & Ives, Z. (2007). Dbpedia: A nucleus for a web of open data. In *The semantic web* (pp. 722–735). Springer Berlin Heidelberg.

오늘날 범용적으로 사용되고 있는 공공데이터라는 말은 그 생산 주체인 ‘정부’, 열린 데이터가 갖는 ‘사용’, ‘재사용’, ‘공유’, ‘재배포’의 속성을 내재적으로 포함하여 인식되고 있다. 각각의 개념이 갖는 차이는 분명하지만 실 데이터 소비자와 데이터를 기반으로 하는 산업계에서의 공공 데이터가 갖는 개념과 위치는 열린 정부 데이터로 보는 것이 바람직하다.

3. 공공데이터 동향

가. 정부 데이터 개방 흐름

최근에는 열린 정부가 새로운 정부 운영 패러다임으로 자리매김하고 있다. 미국, 호주, 유럽연합 등의 개방형 정부 선언은 정부 데이터와 공공데이터 개방을 통해 정부의 투명성 강화와, 차세대 성장 동력인 데이터를 기반으로 하는 경제 체제를 마련해 가는 것이 목표이다.

우리나라는 세계 최고 수준의 정보 통신 기술 기반의 정부 운영, 정보 공유 체계를 구성하여 2010년, 2012년 UN 전자 정부 평가에서 연속 1위를 달성하였지만 미국, 영국 등과 비교했을 때 데이터 개방, 열린 정부 정책 추진에 있어서는 후발 주자로 볼 수 있다.

개별 국가의 데이터 개방을 넘어 국가 간 공동 데이터 개방 협력 추진이 이루어지고 있는 단계로 대표적인 것이 유럽연합의 Open Data Strategy 발표(2011), OECD의 Open Government Partnership(2011), G8 정상 개방형 데이터 협정(Open Data Charter, 2013) 등이다.

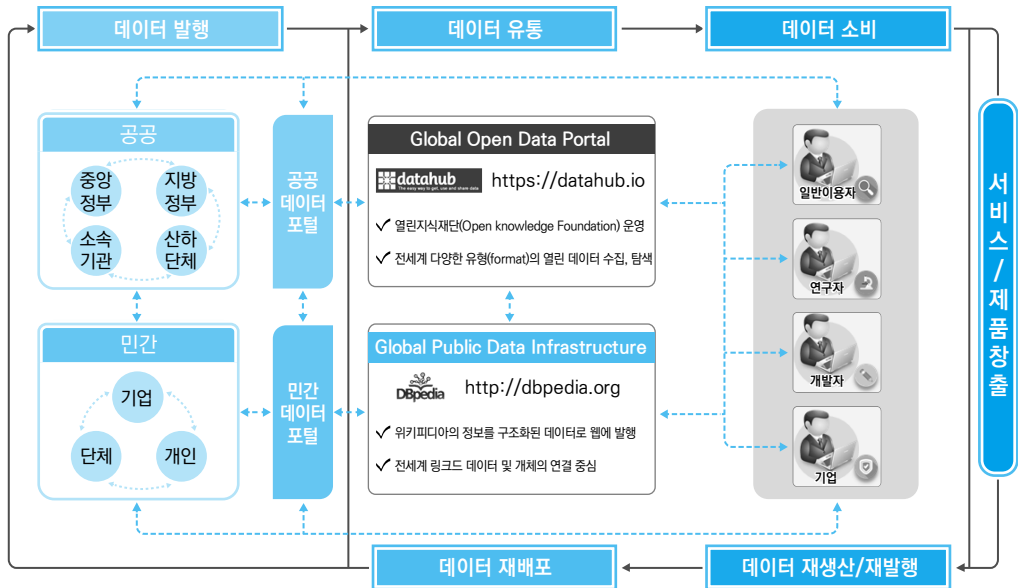
우리나라는 열린 정부 대신 ‘정부 3.0’이라는 용어로 데이터 개방을 추진하고 있는 반면 해외의 경우는 열린 정부 혹은 정부 2.0 등의 용어를 혼용하여 사용하고 있다. 세계적인 컨설팅 그룹 가트너는 정부 2.0과 정부 3.0의 개념을 구분하고 있는데 정부 3.0(Gov 3.0)은 LOD의 결과물인 RDF(Resource Description Framework) 형태로 데이터를 개방하는 단계로 정의하고 있다.

- Gov 2.0: 정보 기술을 활용하여 정부의 서비스, 데이터, 운영을 최적화함
- Gov 3.0: 시맨틱 웹 표준(RDF)를 활용해서 개방형 데이터를 발행

이렇듯 개방형 데이터 생태계에서 정부 데이터, 공공데이터의 개발과 보급의 중요성이 강조되고 각국 정부의 데이터 개방 선언이 이어지면서 초기 웹을 플랫폼으로한 새로운 데이터 생태계에 양질의 공공데이터가 중요한 시장으로 부각되었다. 기존의 정부 정보, 데이터 개방도 구조화된 기계가독형 데이터 공급으로 흐름이 전환되고 있다.

나. 공공데이터 유통 흐름

그림 3-1-2 공공데이터 유통 흐름



기존의 정보 유통과 다르게 데이터 유통이 갖는 특징은 데이터의 재생산, 재발행, 배포가 가능하다는 점이다. 공공기관을 통해 발행된 공공데이터는 국내 공공데이터포털, 민간포털³⁸⁾ 등을 통해서 유통되고, 이들 데이터는 다시 데이터허브(datahub) 등 글로벌 디렉토리에 등재되어 전세계 모든 데이터 이용자를 대상으로 유통, 확산된다. 이용자들은 입수한 데이터를 다시 재가공하여 데이터 유통채널에 재배포하여 데이터 품질은 데이터 생태계 내에서 자발적으로 향상되는 효과를 볼 수 있다.

[그림 3-1-2]에서와 같이 모든 데이터 발행기관과 데이터들이 상호간에 연결된 순환구조가 가능한 이유는 현재의 공공데이터는 기계 가독형 데이터이자 상호간의 인터링크를 기반³⁹⁾으로 하고 있기 때문이다. 일반적으로 이런 유형의 데이터를 링크드 데이터로 볼 수 있으며, 공공데이터의 활성화, 고품질화는 연결 중심의 기계 가독형 데이터 유통에 달려 있다고 할 수 있다.

38) 여기에서의 민간 포털은 네이버, 다음과 같은 기존의 정보 유통 채널이 아니라 기계 가독형 데이터를 유통하는 포털을 의미하며, 한국 의 경우 DBpedia 및 LOD 클라우드에 등재된 KData(www.kdata.kr)를 의미한다.

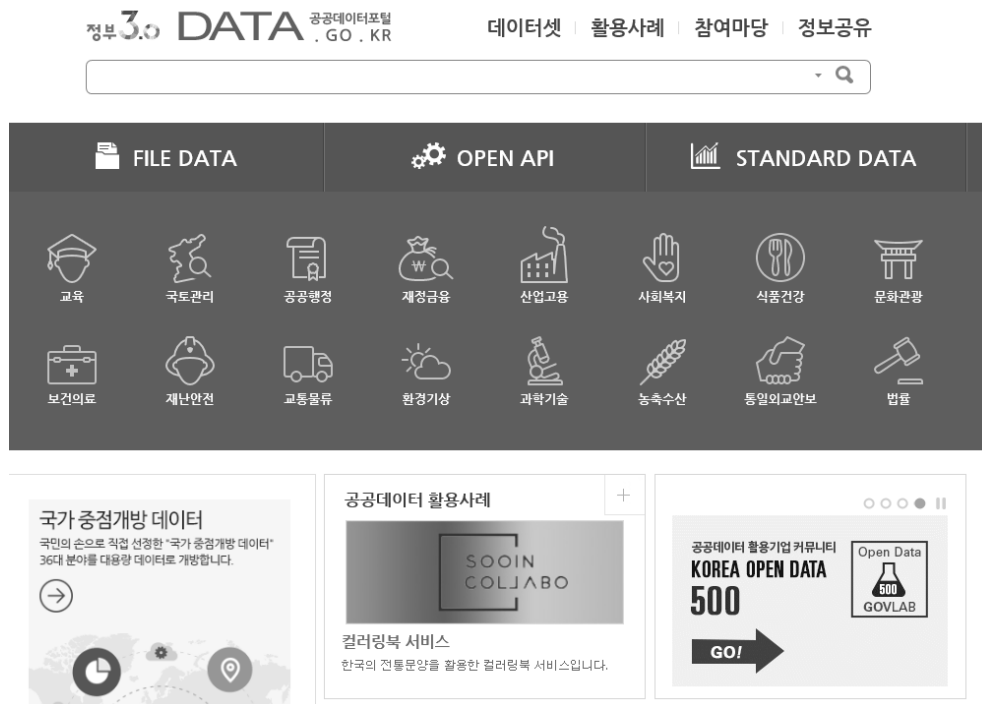
39) 열린 데이터 관점에서 최고 수준의 데이터는 링크드 데이터로 관련된 등급과 데이터 상태는 <http://5stardata.info/ko/>에서 확인할 수 있다.

다. 국내 공공데이터 유통

국내의 경우 공공데이터 개방은 주요 선진국과 비교했을 때 비교적 후발주자에 속한다. 우리나라 정부는 이를 극복하기 위해 2013년 공공데이터의 제공 및 이용 활성화에 관한 법률을 제정하고 이후 단기간 내 데이터 개방을 대폭 확대하였다. 또한 민간 활용도가 높은 핵심 데이터셋을 선정하여 2014년 이후 지속적으로 개방하려는 노력을 지속하고 있다.

- 2014년 이후 공공데이터셋 4,579개 및 오픈 API 143개 신규 제공
- 2014년 이후 민간 활용 핵심 데이터셋 504개 선정 및 우선 개방 추진⁴⁰⁾

그림 3-1-3 공공데이터 포털(<http://data.go.kr>)



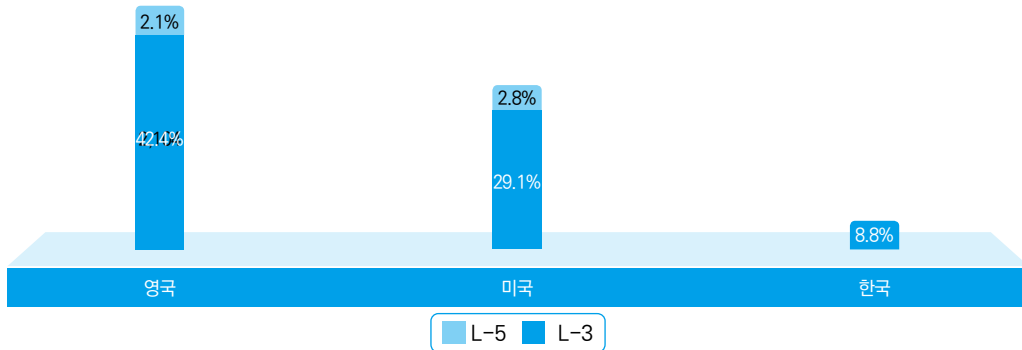
[그림 3-1-3]은 한국정보화진흥원이 운영하고 있는 공공데이터 포털로 전통적인 파일 형태 데이터와 가장 많은 활용성을 갖고 있는 오픈 API 외에 원천 데이터를 표준에 맞게 기계 가독형 데이터로 제공하는 표준 데이터와 함께 제공하고 있다. 표준 데이터 제공과 국가 중점 개방 데이터, Korea Open Data 500은 민간의 활발한 활용을 위한 노력으로 볼 수 있다.

⁴⁰⁾ 관계부처 합동, 2014, 오픈 데이터 5대 강국 도약을 위한 공공데이터 개방 발전전략

라. 국내 공공데이터의 한계

국내 공공데이터가 갖는 한계는 양적 성장에 비해 낮은 활용성을 들 수 있다.

그림 3-1-4 영국, 미국, 한국의 공공데이터 개방 품질 비교(5-스타 오픈 데이터 기준 3단계 이상 비교)



※ 출처: 관계부처 합동, 2014, 오픈 데이터 5대 강국 도약을 위한 공공데이터개방 발전 전략

세계적으로 개방형 데이터의 품질 측정에 가장 범용적으로 활용되고 있는 기준은 5 Star 오픈 데이터로, 개방형 데이터의 품질을 5개 별점으로 단계적으로 구분한다. 별점 3점은 CSV, XML 등과 같이 플랫폼, 소프트웨어 독립적인 기계 가독형 구조적 데이터로 데이터산업 현장에서 직접 활용 가능한 고품질 데이터의 시작으로 볼 수 있다. [그림 3-1-4]에서 보듯이 우리나라의 경우는 공공데이터 개방 활용의 선진국인 영국, 미국에 비해 현저히 낮은 수준을 보이고 있다.

그러나 최근 집중적으로 노력하여 세계 주요국의 데이터 개방 노력과 비교했을 때 현저하게 발전한 것을 확인할 수 있는데, 오픈 데이터 바로미터(Open Data Barometer(<http://opendata-barometer.org>))가 제공하는 지표 결과가 이를 방증한다.

표 3-1-1 오픈 데이터 바로미터 한국 순위

구 분	전체 순위	점수	준비성	실행력	영향력
2015	8	71.19	95	64	58
2014	17	57.65	79	54	48
2013	12	54.21	77	55	25

우리나라는 조사 첫 해 전세계 12위에서 2014년 17위로 선진국에 비해 낮은 수준을 유지하고 있었으나 최근 조사에서는 8위를 기록했다. [표 3-1-1]의 세부 점수에서 알 수 있듯이 준비성 부분에서 가장 높은 성장세를 보이고 있는데 이는 정부의 공공데이터 개방을 위한 정책 추진, 양적 확장을 위한 노력의 결과로 볼 수 있다. 그럼에도 시장에서 실제 활용을 가능하게 하는 영향력 부분은 여전히 낮은 점수를 유지하고 있다.

4. 경제적 효과와 최신 기술 동향

가. 경제적 효과

공공데이터에 대한 경제적 효과 전망은 열린 데이터 측면에서의 접근이 일반적이다. 유럽연합은 열린 데이터에 대한 경제적 효과를 연간 40조유로(약 5,632조원)로 예측⁴¹⁾하고 있으며, 데이터 경제 활성화를 위한 연구개발을 위해 공동 자금을 지원하고 있다. 대표적인 프로젝트는 아래와 같다.

- LOD2(<http://lod2.eu/>): 2010년 9월부터 4년간 수행. 데이터를 보다 유용하게 만들 수 있는 시맨틱 기술들을 활용한 링크드 데이터 생명주기 연구와 핵심 기술(제품) 활용 및 개발
- OpenAIRE(<http://www.openaire.eu/>): 2009년 12월에 시작된 프로젝트로 유럽연합 25개국과 관련 국가가 참여하여 연구 정보 개방형 접근을 위한 기반 구조 구성
- ISA조치(<http://www.semic.eu/>): 각 국가는 물론 유럽연합 수준에서 메타데이터 정렬을 위한 첫 번째 단계로 개방형 정부 메타데이터의 의미적 상호 운영성 확보를 위한 연구

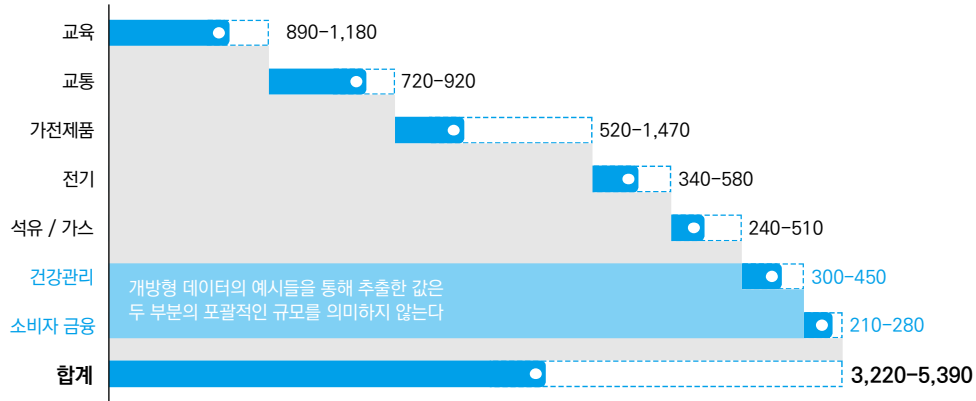
맥킨지(McKinsey)⁴²⁾ 보고서에 따르면 개방형 데이터의 효과로 7개 주요 부문에서 연간 3조달러(약 3,000조원) 이상의 경제적 부가가치가 증가할 것으로 예측하고 있다. 7개 주요 분야는 교육, 교통, 가전제품, 전기, 석유/가스, 건강 관리, 소비자 금융으로 세부 내용은 [그림 3-1-5]와 같다.

41) Open data – An engine for innovation, growth and transparent governance
<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0882:FIN:EN:PDF>

42) Open data: Unlocking innovation and performance with liquid information
<http://bit.ly/11FQr3H>

그림 3-1-5 개방형 데이터의 잠재적 가치

(단위 : 10억달러)



※ 출처: Open data—Unlocking innovation and performance with liquid information, McKinsey

글로벌 컨설팅사 캡제미니(Capgemini)는 개방형 데이터가 갖는 경제적 효과에 대해 정부 차원과 민간 차원으로 구분하여 수익 창출, 비용 절감과 효율성 유도, 고용 창출과 미래 지향적 기술 개발 부분을 지적하였다.

표 3-1-2 개방형 데이터의 경제적 효과

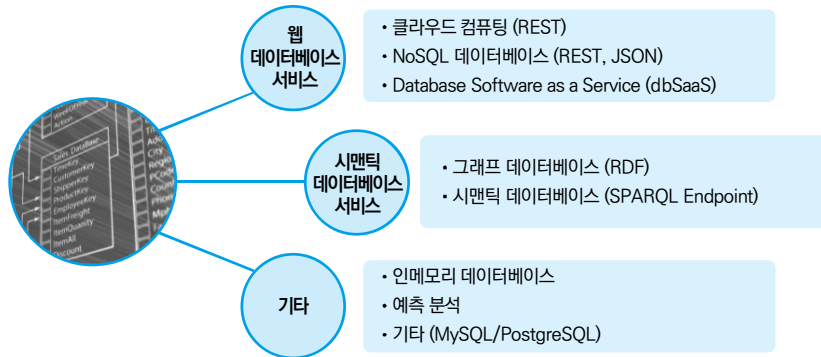
	다양한 분야에서 수익 창출	비용 절감과 효율성 유도	고용 창출과 미래 지향적 기술 개발
정부 차원의 혜택	- 경제활동 증가로 인한 세입 증가 - 고부가가치 정보 가격이 반영됨으로써 수익 증대 가능	- 거래 비용 감소 - 링크드 데이터를 통한 서비스 효율성 증가	- 일자리 창출 - 기업가 정신 고취
민간 차원의 혜택	새로운 비즈니스 기회 창출	- 원천 정부 데이터에 대한 변환 투자가 없으므로 비용 절감 가능 - 정확한 정보에 기반한 보다 나은 의사결정 가능	숙련된 인력 확보

※ 출처: The Open Data Economy, 캡제미니 컨설팅, 2013

나. 기술 동향

한국정보화진흥원은 2014년 전세계 공공데이터 현황조사를 통해 데이터베이스 활용 기술 동향을 분석하고 공공데이터, 데이터베이스 분야에서 주목해야 하는 기술 동향을 파악하였다.

그림 3-1-6 최신 데이터베이스 기술 동향



※ 출처: 데이터베이스 활용기술 전망, 한국정보화진흥원, 2014

보고서에서는 웹 데이터베이스 서비스 기술, 시맨틱 데이터베이스 서비스 기술 분야가 최신 기술로 인식되고 있음을 지적하였다. 또한 국내 공공데이터의 경우 웹 데이터베이스 서비스, 지리 데이터베이스 서비스, 시맨틱 데이터베이스 서비스 순으로 중점적인 기술 투자와 데이터 구축이 필요함을 언급하였다. 대표적인 기술 형식으로 웹 데이터베이스의 경우는 REST, JSON, 지리 데이터베이스의 경우 WMS, KML, 마지막으로 시맨틱 데이터베이스의 경우 SPARQL Endpoint, RDF를 들었다.

5. 향후 전망

오늘날 공공데이터는 단순히 공공 기관에 의해서 생성, 서비스되는 데이터 이상의 의미를 갖는다. 공공데이터는 새로운 경제 질서의 중심에 있으며, 정부는 단순한 데이터의 제공자가 아니라 고품질 데이터를 민간 산업계에 공급해야 하는 중요한 임무를 갖는다. 공공데이터의 중요성이 높아지게 된 기술적·사회적 요구는 다음과 같이 요약할 수 있다.

- 가공 전 원천 데이터 공급의 필요성 증대
- 재활용이 가능한 라이선스, 표준 준수 데이터의 필요성 증대
- 기계 가독형 구조적 데이터 중심의 생태계 등장
- 웹을 플랫폼으로 한 데이터 생태계의 등장

이런 요구들은 향후 공공데이터 자체의 수요, 공급은 물론 공공데이터 활성화와 민간 활용성 증대를 가속화할 수 있는 기술 시장에도 반드시 수용해야하는 요소들이다.

향후에도 정부, 공공기관의 책임있는 데이터 생산, 공급 체계와 더불어 표준화된 데이터 품질 체계를 통해 활용성이 극대화된 공공데이터 시장 체계 구축이 필요하다.

제2절 기상정보

1. 기상정보 개요

가. 기상정보의 정의

기상정보는 현행 법령상 용어 정의가 명시되어 있지 않으나, 기상청 기상백과에서 ‘기상에 관한 지식이나 보도 내용으로서, 일기 예보·기상 통보·기상 주의보·경보 등으로 정의되며 좁은 의미로는 기상관서가 일기 예보 등과 관계있는 태풍·폭풍이나 그 밖의 악기상 등에 대한 정보를 일반 국민이나 관계 기관에 구체적으로 신속히 발표하는 것’으로 정의된다. 또한 기상청훈령인 예보업무규정 제40조(기상정보)에는 ‘현재 및 예상되는 기상 상황을 국민에게 특별히 알려주어야 할 필요가 있을 때 행하는 기상 현상에 대한 설명과 특보 이외의 예보’로 정의하고 있다.

그러나 이는 기상 상황 또는 기상 현상에 대한 상세한 정의를 포함하고 있지 않으며, 정보 제공 주체 역시 맥락상 기상청으로 한정되어 있어 현재 다양한 제공 주체가 존재하는 기상정보를 정의하기에는 한계가 있다. 따라서 ‘기상’과 ‘정보’의 합성어인 기상정보의 경우, 기상 현상에 대한 정의를 통해 총체적으로 도출될 수 있다. 기상 현상이란 기상법 제2조(정의)에 의하면 대기의 여러 현상을 의미하는 기상(氣象)과 더불어 기상에 영향을 미치는 현상인 지상(地象; 지면 또는 지중에서 일어나는 현상)과 수상(水象; 내륙의 하천과 호수 또는 해양에서 일어나는 현상), 기타 대기권 밖의 현상을 모두 포함한다. 따라서 기상정보는 기상, 지상 및 수상 등에 관한 과거, 현재 및 미래의 정보라고 종합적으로 정의할 수 있다⁴³⁾.

나. 기상정보의 활용

기상정보의 활용은 목적별로 분류될 수 있다. 첫 번째, 방재 정보로서의 기상정보 활용이다. 기상 예보 및 특보 정보 등을 국가 및 개인 등이 방재 활동에 활용하여 태풍, 대설, 집중호우 등의 기상재해로부터 인명과 재산을 보호하는 것을 목적으로 한다.

두 번째, 국민 삶의 질을 높이는 정보로서의 기상정보 활용이다. 인명 및 재산 등에 대한 직접적인 피해는 작으나 더위·추위, 강수, 바람 등 생활 측면에서 불편을 야기할 수 있는 기상 현상에 대하여 의상 선택과 같은 일상생활과 각종 레저 활동 등에 참고하여 생활 편의를 누리거나 또는 그 수준을 향상시키는 것을 목적으로 한다.

세 번째, 안보 정보로서의 기상정보이다. 각종 군사작전에 대한 기상정보 활용으로 안보 역량을 높이거나 물관리, 에너지관리 등 자원 안보 측면에서도 기상정보가 활용된다.

네 번째, 산업계 경영 정보로서의 기상정보 활용이다. 앞서 제시한 세 가지 부문이 전통적인

43) 시그마프레스(2013), 기상정보론

기상정보 활용 부분이었다면, 이는 기상정보를 이윤 창출과 경영 효율의 수단으로 활용하는 적극적이고 능동적인 활용 방식이다.

최근 기상정보는 지역적 세밀성과 정보의 다양성을 기반으로 점차 세분화·고도화 되어가는 산업계에서 경영 환경을 효율화하고 부가가치를 창출하기 위한 방안으로 인식되고 있다. 특히 기후 변화로 인한 이상 기상 현상의 빈도와 강도가 증가되면서 산업별로 긍정적/부정적 요소로 작용할 수 있는 기상·기후 리스크의 체계적 관리를 위해 ‘날씨 경영’을 적극 도입·활용하고 있다. 이는 날씨가 제품 생산 공정에 영향을 주거나, 소비자의 소비 심리에 영향을 미쳐 제품의 품질 제고와 매출 증대로 연결되기 때문이다.

날씨 경영이란 생산, 기획, 마케팅, 영업 등 기업 경영의 다양한 분야에 날씨를 적용하여 기업의 이윤 창출 및 경영 효율 증대를 도모하는 것으로, 미국, 일본 등 선진국에서는 오래전부터 날씨를 유가나 환율 급리처럼 중요한 경영 변수로 인식해 적극 활용 중이다⁴⁴⁾. 최근 국내에서도 다양한 산업에서 날씨 경영을 활용하여 우수한 성과를 창출한 사례들이 보고되고 있다.

표 3-1-3 업종별 날씨 영향 및 날씨 경영 사례

업종	날씨 영향 ⁴⁵⁾	날씨 경영 사례
농업	• 수해, 가뭄, 이상 고온 등에 따른 생산 변동 • 병충해 및 품질에 영향	• J사는 벼 생산 및 수확에 맞는 특화된 기상정보 활용으로 적기 수확에 성공, 연간 9억원의 부가가치 창출
건설업	• 공사 현장 재해, 공정 관리, 품질관리, 안전 관리, 원가관리 등에 영향	• L사는 타워크레인에 풍속계와 자가우량계 설치, 기상정보 활용을 통한 일정 관리 등 날씨 경영 활동으로 피해 복구 비용 연간 약 910억원 절감 효과
운송업	• 항공기·선박 결항, 교통체증 등으로 운송 비용 및 운송 시간에 영향	• A사는 기상정보 활용으로 항공기 회항 31% 감소하여 비용 절감 효과, 기상정보 기반 비행 감시 시스템 활용으로 각종 비정상 기상 상황에 신속히 대처
레저업	• 여름 이상 저온, 이상 다우 시 레저객 감소 • 겨울 이상 고온 시 스키장 고객 감소 및 운영 비용 증가	• J사는 고객 맞춤형 기상정보 서비스로 박람회 방문객 1인당 매출액 23% 추가 증가
유통업	• 폭염·혹한 시 백화점 매출 감소, 인터넷 쇼핑, 홈쇼핑 매출 증가	• C사는 날씨정보시스템 운영으로 물류센터 재고 일수 50% 감축, 매출 30% 증가, 폐기량 40% 감소
외식업	• 식재료 관리, 편차가 심한 배달 수요, 원가 관리 등에 영향	• V사는 기상정보를 고려한 고객수 및 수요 예측을 통해 식자재 폐기 비용 10% 절감
에너지업	• 여름 저온, 겨울 고온 시 전력·가스 생산업 매출 감소 • 가뭄 시 수력 발전의 발전량 감소 • 폭설 시 전력 송전 이상 발생	• K사는 기상청 기상 통계 자료 활용을 통한 해상 장비 사용 효율 28% 증가로 약 5억원의 발전소 건설 비용 절감, 기상 재해로 인한 발전 설비 고장 사고 0건 달성
제조업	• 여름 이상 저온 시 음료, 주류, 에어컨 제조업 매출 감소 • 겨울 이상 고온 시 패션업계 계절상품 재고 물량 증가, 매출 감소 • 황사 시 기계조립, 초정밀산업 생산 비용 증가	• 식품업 P사는 제품별 날씨 판매 지수를 마케팅에 활용하여 판매량 예측, 주문량 조절, 재고 폐기 물량 감소 • 의류업 B사는 날씨에 따른 제품 수요 예측 솔루션을 개발하여 전년 대비 323억원 매출 증가, 불철 재킷 매출 60% 증가

⁴⁴⁾ 기상청(2013), 기상정보 활용 저변 확대를 위한 날씨 경영 인증제도 활성화 지원 연구

⁴⁵⁾ 기상청(2013), 기상정보 활용 저변 확대를 위한 날씨 경영 인증제도 활성화 지원 연구

2. 국내 서비스 현황

가. 공공 서비스 현황

1) 대국민 서비스 : 기상자료개방포털

기상청은 공공데이터의 민간 이용 활성화를 통한 데이터의 가치 제고를 목표로, 국민이 기상 자료를 쉽게 찾아 이용할 수 있도록 기상 자료 서비스 통합 창구인 기상자료개방포털을 운영 중이다. 지상·해양·고층·역사 기후 자료 등 9개 부분 총 25종의 데이터와 간행물 정보를 다운로드 방식으로 제공한다.

표 3-1-4 기상자료개방포털 제공 데이터 목록

구 분	제공 데이터
기상 관측 자료(4종)	지상에서 자동관측 또는 목적(目測)한 종관, 방재, 농업, 북한 기상 관측 자료 * 기온, 기압, 강수량, 풍향, 풍속, 운량, 시정, 적설, 일조, 일사 등
해양 관측 자료(3종)	해상에 설치된 부이, 등표, 파고부이 관측 자료 * 수온, 파고, 기온, 기압, 풍향, 풍속 등
고층 관측 자료(2종)	지상에서 약 30km 상공까지 연속으로 관측한 레원존데, 연직바람관측(윈드프로파일러) 자료 * 기압, 기온, 습도, 풍향, 풍속 등
황사 관측 자료(1종)	부유 분진 측정기에서 관측한 미세먼지(PM10) 자료
레이더 자료(2종)	기상 레이더에서 관측한 지점별 자료, 전체 레이더 합성 자료
위성 자료(1종)	우리나라의 정지 궤도 위성인 천리안 위성 기상 관측 자료
수치 예보 자료(4종)	수치 예보 모델인 지역(RDAPS), 국지(LDAPS), 초단기(KLAPS), 지역파랑 모델 자료
기상 자원 지도(2종)	풍력과 태양에너지 분포를 나타내는 풍력기상자원지도, 태양기상자원지도
역사 기후 자료(6종)	조선왕조실록 및 각사 등록에서 발췌한 기상 기록과 해관자료, 일기도, 자기기록지, 통계원부류 등 과거 종이 기록물을 스캔하여 복원한 자료
간행물	기상청에서 발간하는 기후통계 간행물인 기상월보·연보, 기후표 등을 링크함

또한 오픈 API 기능을 제공하여 6종의 데이터⁴⁶⁾에 대해 자료 생산 즉시 API로 서비스한다. 특히 데이터 검색시 사용자가 필요한 지점에 인접한 관측 지점의 정보 조회가 가능함에 따라 최적의 관측데이터 확보를 지원한다. 데이터 제공 외에도 기후통계분석 기능⁴⁷⁾을 통해 사용자가 원하는 데이터의 통계분석 결과를 도표, 그래프, 이미지 등으로 표출·제공함으로써, 데이터 분석 서비스를 제공한다.

46) 오픈 API 서비스 데이터(6종) : 동네예보, 중기예보, 예보구역, 생활기상지수, 보건기상지수, 항공기상정보

47) 기후통계분석 서비스 : 기온분석, 강수량분석, 극값순위, 기후평년값, 장마, 황사, 폭염일수, 열대야일수, 24절기의 기후분석, 체감온도, 실효습도, 열지수, 냉·난방도일, 적산온도

그림 3-1-7 기상자료개방포털 서비스



※ 출처: 기상자료개방포털 홈페이지(<http://data.kma.go.kr>)

2) 유관기관 대상 서비스 : 클라우드 방재기상정보시스템

‘클라우드 방재기상정보시스템’은 기상 빅데이터와 기상 분석 기술을 적용한 클라우드 환경에서 기상정보를 공동으로 활용할 수 있는 시스템으로, 기상정보를 활용하는 다양한 기관(군, 방재, 산림, 수문 등)에 특보 등 기상정보를 공유하고 정보의 검색·분석을 지원함으로써 방재부문에서의 정부 기관간 공유·협력 체계를 지원한다.

3) 기상사업자 대상 서비스 : 기상정보 제공

기상청과 한국기상산업진흥원(이하 ‘진흥원’)은 기상산업진흥법 제13조(기상정보의 활용 촉진 등) 및 동법 시행규칙 제7조(기상정보의 제공 방법 및 수수료 등)에 의거, 등록된 민간 기상사업자에 기상정보를 제공하고 있다. 이는 기상청에서 생산된 기상정보를 진흥원의 전용 서버를 통해

////////////////////////////////////

표 3-1-5 기상사업자 제공 기상정보

※ 출처: 한국기상산업진흥원 홈페이지(<http://www.kmipa.or.kr>)

그림 3-1-8 기상정보 제공 흐름도



나. 민간 서비스 현황

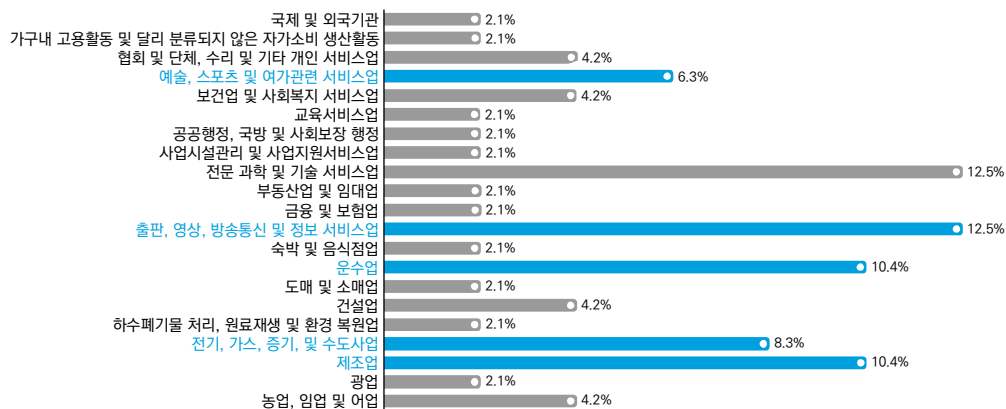
민간 기상 기업은 기상청 등으로부터 제공받은 기상정보를 재분석하거나 타 분야 정보 등을 융합하여 기상정보 서비스 비즈니스 모델을 창출하고 있다. 현재 국내 기상 서비스 기업은 39개사⁴⁸⁾로 집계(기상청 등록기업 기준)되고 있으며, 기상예보·컨설팅·감정 서비스 등을 제공한다.

표 3-1-6 민간 기상서비스(기상산업진흥법 제2조 관련)

구 분	내 용
기상 예보	기상 현상에 관하여 관측된 결과를 바탕으로 미래의 기상 상태를 예상하여 제공
기상 컨설팅	기상정보를 분석·평가하여 경영 활동에 관한 조언을 제공
기상 감정	기상 현상에 관하여 관측된 결과를 바탕으로 특정 지점의 기상 현상을 추정하거나 그 기상 현상이 특정 사건에 미친 영향의 정도 등을 판단

기상 기업의 경우 산업계 고객을 주 서비스 대상으로 하며, 연구 결과⁴⁹⁾⁵⁰⁾에 의하면 ‘출판, 영상, 방송통신 및 정보 서비스업(방송, 미디어)’, ‘운수업’, ‘제조업’, ‘전기, 가스, 증기 및 수도사업(에너지, 수자원)’, ‘예술, 스포츠 및 여가관련 서비스업(관광·레저)’에 대한 서비스 비율이 높게 나타났다.

그림 3-1-9 민간 기상 서비스 공급업종 현황



※ 출처: 업종별 기상정보 활용 사례조사, 한국기상산업진흥원, 2016

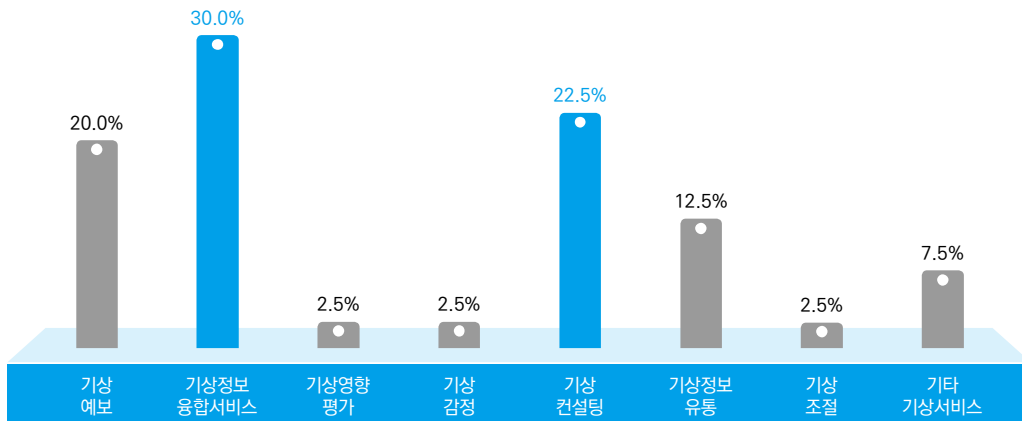
⁴⁸⁾ 2016. 8월 기준 기상청 등록 기상기업은 350개사이며, 기상정보업만을 등록한 기업(311개사)을 제외한 기상예보업, 컨설팅업, 감정업 중 하나라도 등록한 기업 수를 집계

⁴⁹⁾ '전문, 과학 및 기술 서비스업'의 경우, 기상기업의 R&D 및 연구과제 수행 등에 따른 용역 서비스로 분류됨에 따라 제외

⁵⁰⁾ 한국기상산업진흥원(2016), 업종별 기상정보 활용 사례조사

또한 주요 기상 서비스 제공 현황의 경우 에너지, 건설, 환경 등 타분야 정보와 융합한 ‘기상정보융합 서비스’ 및 ‘기상 컨설팅’ 제공 비율이 높은 것으로 나타나, 산업별·고객별 맞춤형 서비스 부문에 주력하고 있는 것으로 나타났다.

그림 3-1-10 민간 기상 서비스 공급업종 현황



※ 출처: 업종별 기상정보 활용 사례조사, 한국기상산업진흥원, 2016

3. 해외 서비스 현황⁵¹⁾

가. 미국

1) 공공 기상 서비스 현황

미국해양대기청(NOAA)의 산하기관인 미국기상청(NWS)이 기상정보 자체 홈페이지 및 라디오 등을 통해 전달하나, 공공부문과 민간부문의 갈등을 피하고 협력을 강조하는 민관 협력 방침에 따라 기상청이 기상정보를 직접 유통하기보다는 기상사업자가 언론이나 상업적 기업 또는 최종 수요자인 국민에게 전달하는 비중이 크다.

2) 민간 기상 서비스 현황

미국의 민간 기상 서비스 분야로는 기상 장비(원격 탐사), 일기 예보 서비스, 인터넷과 모바일 기기를 이용한 기상정보 제공, 산업 기업 내의 기상 전문 직원, 기상 컨설팅 서비스 및 기상과 기후

51) 기상정보 민간활용 지원체계 구축 및 기상산업진흥기본계획 수립, 한국기상산업진흥원, 2015

연구/조사, 법의학 기상학, 전문 기업 서비스(기상 조절, 번개 감지기, 수직 측풍 장비, 생명기상학 등), 환경 컨설팅 및 엔지니어링 대기질 감시, 기상시스템 개발, 미디어 기상학(지역 TV 기상학자 포함), 기상 위험 관리, 기상 교육 등으로 다양한 서비스가 제공되고 있다⁵²⁾.

나. 일본

1) 공공 기상 서비스

일본의 경우 방재(기상 및 지진 특보 등) 및 R&D 업무를 중심으로 수행되고 있으며, 이외 대부분의 기상 서비스는 예보 업무 허가를 받은 민간 기상사업자의 영역이다⁵³⁾. 기상정보의 전달은 주로 방송, 웹, 방재용 등으로 이루어지고 있으며, 방송의 경우 이미지와 그림 위주로 전달이 되고, 방재 기관은 계정을 통해 시스템에 접속하여 정보를 입수한다. 기상사업자에 대한 서비스의 경우, 기상청 산하의 기상업무지원센터가 기상청으로부터 수신한 각종 기상 데이터를 기상사업자들에게 송신하고 있으며, 과거 관측 및 통계 데이터의 경우 CD 등의 매체로 제공한다.

2) 민간 기상 서비스

예보 업무 허가 사업자는 휴대전화, 인터넷 등을 통해 지역적·시간적으로 세밀한 기상정보를 제공하거나 각 고객기업의 위험 관리 서비스 및 개인의 요구에 대한 특화된 기상정보 등의 서비스를 제공한다. 일본에서는 국내와 유사하게 민간 예보업무에 대해 허가제로 운영하고 있으며(단, 기상장비업은 제외), 기상/파랑 예보 업무 허가를 받은 민간 기상 사업자는 122개로(2015년 기준) 조사되고 있다.

다. 영국

영국은 정부와 민간이 완전 경쟁 체계를 형성하여 서비스를 제공하고 있으며, 이를 통해 시장을 활성화하고 있다. 기상 서비스의 종류나 형태에 따라 무료/유료가 구분되며 날씨 예보 및 재해 기상예보 등의 무료서비스는 주로 대중매체 및 인터넷으로 제공하고, 각종 예보관 상담서비스(레저, 에너지, 금융·보험, 교통 등)는 전화 및 팩스 등을 통해 유료로 제공하고 있다.

⁵²⁾ 기상정보 유통 및 활용 기술 연구회, (재)기상지진기술개발사업단, 2010

⁵³⁾ 선진국(일본) 기상산업 현황 및 민간 기상산업 육성 정책조사 공무국외여행보고서, 한국기상산업진흥원, 2015

4. 향후 전망

최근 기상청은 택시의 운행 기록 자기진단장치(OBD)에서 수집한 기온 및 기압 등 기상 데이터를 관측 자료로 활용하기 위해 기술개발을 추진 중에 있다. 이는 기상관측소가 설치되어 있지 않은 관측 공백 공간에 대한 밀도 높은 기상데이터의 수집을 시도하는 것으로, 해당 기술이 성공적으로 개발될 경우 기존 13km 간격으로 설치된 기상관측소의 위치적·공간적 한계를 해소하게 될 것이다. 이에 따라 고밀도의 정확도 높은 예측 정보 생산이 가능해져 기상정보에 대한 활용도 및 국민 체감도 향상에 기여할 것으로 보인다.

미국, 영국 등의 경우에도 개인용 기상 관측 장비 보급, 다양한 디바이스 정보 및 웹데이터의 수집, 택시 등 운송수단을 통해 다량의 기상 관측 데이터를 수집하여 활용을 추진 중이다. 이는 컴퓨팅파워의 확장과 함께 빅데이터 처리 기술의 발달과 연결되어 있으며, 결국 이러한 기상정보는 빅데이터플랫폼을 통해 오픈됨으로써 범죄, 교통 등의 사회 현안 해결 및 경제적 인사이트 도출에 활용될 것으로 전망되고 있다.

국내에서도 드론 등과 결합된 IoT 기상관측센서 등을 통해 기상·기후 빅데이터의 종류와 용량이 폭발적으로 증가될 것으로 기대되며, 이러한 시도는 기상정보의 정확도 향상과 더불어 다양한 산업 정보와 융합되어 서비스됨으로써 창조경제 실현의 원동력이 될 것으로 전망된다.

제3절 보건의료정보

1. 보건의료정보 개요

과거 입에서 입으로 전달 되던 정보는 문자가 개발되면서 비약적으로 확대되어 유형의 기록으로 남게 되었고, 이제는 손에 들고 있는 스마트폰 하나로 전세계 모든 곳에서 생산된 정보를 실시간으로 볼 수 있는 시대가 되었다.

특히 보건의료정보는 정보가 가진 전문성으로 인해 정보를 생산하는 전문가 집단과 소비자 사이의 간극이 가장 큰 분야 중 하나이면서 또 대부분의 나라에서 정부 주도하에 생산되고 관리됨에 따라 공유와 개방이 가장 늦은 분야였다. 그러나 정보를 중개하는 다양한 기업들이 생겨나면서 정보의 폐쇄성이 허물어지며 새로운 서비스들이 등장하기 시작했다.

보건의료정보는 진료 과목별 의료기관 개·폐업 현황과 인구 동태 통계를 연결한 상권 분석, 지역별 처방 실적 데이터를 활용한 제약회사의 마케팅 전략 수립 등 B2B에서 활용하는 보건의료산업 데이터와 내 주변 지역의 좋은 병원 검색, 증상으로 질병 가능성을 분석해 주는 인공지능 등 B2C 고객의 건강 증진용 콘텐츠로 구분 가능하다.

2. 국내 서비스 현황

가. 공공 서비스 현황

현 정부 출범 이래 공공 정보를 적극적으로 개방하고 공유하는 정부 3.0 사업이 시작되며 보건의료분야의 정보 개방성도 큰 폭으로 확대되었다.

1) B2C 대상 보건의료정보 제공 현황

보건복지부와 대한의학회가 국민 건강 증진을 위해 바른 건강 정보를 제공할 목적으로 운영중인 국가건강정보포털은 기본적으로 누구나 자유롭게 콘텐츠를 퍼갈 수 있도록 설계했다. 일반 사용자는 원하는 콘텐츠를 선택하고 본인이 운영 중인 블로그, 카페 등에 쉽게 복사할 수 있는 건강 담기 서비스를 이용할 수 있다.

기업 등 기관 사용자는 자사 서비스에 국가건강정보포털 콘텐츠를 서비스할 수 있도록 일부 콘텐츠 또는 전체 콘텐츠의 API를 요청할 수 있다. 국가건강정보포털의 콘텐츠는 매년 업데이트가 되어 건강 정보의 최신성을 항상 유지하고 있다.

이외에도 국립보건연구원이 제공하는 희귀난치성 질환정보, 기상청의 생활 건강 지수, 한국환경공단의 미세먼지 지수 등 국민 생활에 도움되는 정보를 공공기관이 직접 제공하고 있다.



2) B2B 대상 보건의료정보 제공 현황

정부와 지자체는 지역별 의료기관 등록 현황, 음식별 영양소 및 칼로리 정보, 건강보험심사평가원(이하 심평원)에서 진행한 1차 의료기관의 만성질환 환자 관리 평가 결과, 병원별 급여·비급여 항목에 대한 비용 정보, 시도별 및 연령·성별 의약품 사용 현황 등 1,000여 건의 보건의료 기초 정보를 제공하고 있으며, 이 중 971건의 데이터를 공공데이터포털(data.go.kr)을 통해 공개하고 있다. 또한 본 데이터에 대한 접근성을 강화하기 위해 네이버 데이터 랩(datalab.naver.com)에서도 검색 연동이 되어 있다.

정부가 제공하는 데이터는 파일 형태의 데이터셋을 직접 제공하는 방식과 API를 연동하는 형태로 구성되어 있는데, 민간기업들이 공공데이터를 보다 효율적으로 관리, 이용하기 위해 API 제공이 더 확대되어야 할 것으로 보인다.

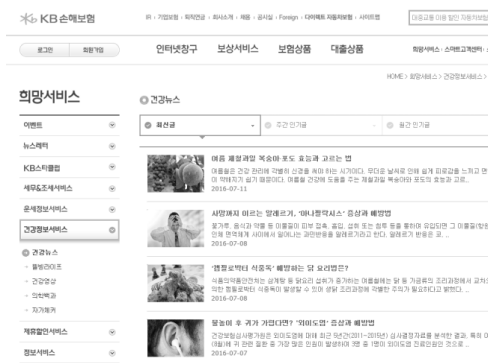
나. 민간 서비스 현황

각 기업들은 사용자 수를 늘리고 수익을 창출하기 위해 자체적으로 개발한 보건의료 정보를 활용하거나 정부에서 공개한 데이터를 재가공하며 매시업 서비스를 발굴해 운영하고 있다.

1) 하이닥(HiDoc)

네이버 지식iN 의료 상담으로 유명한 하이닥 서비스는 네이버, 다음 등 포털에 발행되는 건강 관련 칼럼, 뉴스 등의 서비스를 API 형태로 병원 또는 기업에 제공하고 있다. 10만 건 이상의 주제별 건강 콘텐츠가 등록되어 있고, 매월 700건 이상의 콘텐츠가 신규 추가된다. 현재 3,500여 병의원, 동부화재 등 보험사, 제약사처럼 건강 콘텐츠가 필요한 산업 군에서 하이닥 건강 콘텐츠 API 서비스를 이용 중이다. 하이닥은 예약 플랫폼 서비스도 제공하고 있는데, 사용자가 하이닥에서 특정 병원을 예약할 수 있고, 예약 신청이 등록되면 병원에 알림이 간다. 병원은 하이닥 예약 위젯을 병원 EMR, 홈페이지, 블로그에도 설치할 수 있어 하이닥 예약 서비스만으로도 병원의 고객을 통합 관리할 수 있다.

그림 3-1-11 하이닥 건강 콘텐츠 API가 적용된 보험사와 병원 홈페이지

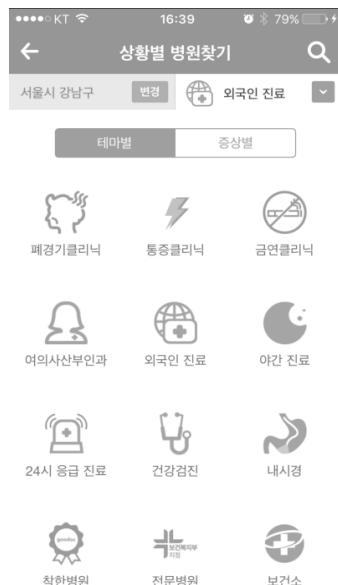


※ 출처: 각 사 해당 홈페이지 캡처

2) 굿닥

모바일 전문 기업 옐로모바일 자회사인 옐로오투오가 운영하는 서비스로 심평원 데이터를 기반으로 접속 지역 중심의 병원, 약국 정보를 소비자에게 제공한다. 심평원 정보 외 병원에서 진행하는 비급여 시·수술 정보도 등록되어 주로 젊은 여성을 타깃으로 병원 예약을 중계하는 이벤트 서비스도 제공된다. 현재 6만 7,000여 병원의 정보, 2만 4,000여 약국 정보, 454개의 이벤트 정보가 등록되어 있다.

그림 3-1-12 굿닥 서비스 이미지



3) 기타

비트컴퓨터가 투자한 것으로 알려진 ‘똑닥’ 서비스는 현재 심평원 병원 정보 제공 중심 서비스로 전자차트 서비스와 연계해 의료기관의 누적된 데이터를 통한 고객 관리 중심 서비스로 진화하고 있다.

삼성전자가 운영 중인 글로벌 모바일 애플리케이션인 S Health는 국내 사용자 기준 월 300만 명이 이용하는 1위 모바일 헬스케어 서비스다(랭키닷컴 기준). S Health는 칼로리, 영양소 등 식품 정보 DB를 최고 수준으로 확보해 제공하고 있으며, 사용자가 활동을 하면서 생성되는 라이프로그 데이터를 누적 제공한다. 현재 자기수치화 데이터를 활용한 가장 앞서 있는 웰니스 서비스로 고객과 보건의료정보를 확보한 서드파트 사업자와 제휴로 추가 서비스를 준비하고 있다.

화해(화장품을 해석하다) 애플리케이션은 국내에서 시판중인 6만 6,000여 화장품의 전성분이 분석되어 있고, 120만건 이상의 사용자 생성 콘텐츠(UGC, User Generated Contents) 상품평이 등록된 화장품 전문 데이터 서비스로 분류된다.

또한 화장품에 사용되는 1만개 이상의 성분들을 백과사전과 같이 위험도에 따라 자세한 내용을 전달해 사용자가 좋은 화장품을 고를 수 있도록 가이드를 제공하고 있다.

3. 해외 서비스 현황

가. 공공 서비스 현황

1) 영국

사회보장제도를 최초로 도입한 영국은 현재 NHS(National Health Service)를 통해 무상의료 서비스를 제공하고 있다. 또한 보건의료 정보를 효율적으로 생산하고, 안전하게 관리하기 위해 정부부처와 독립된 조직인 HSCIC(Health and Social Clinic Information Center)를 설립해 보건의료 관련 빅데이터를 관리하고 있다.

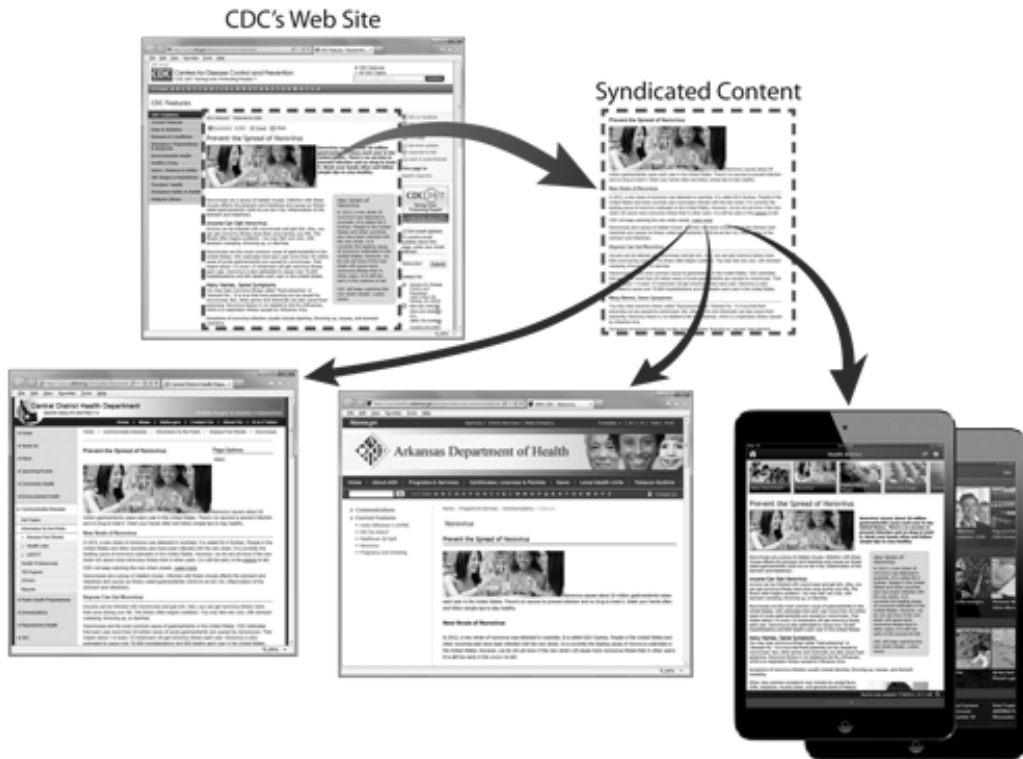
HSCIC는 국민 건강증진을 통해 보험 재정을 효율적으로 관리하기 위해 예방과 질병 관리에 도움이 되는 질 좋은 콘텐츠를 NHS Choices 서비스를 통해 운영하고 있으며, 콘텐츠는 헬스케어 기업들에게도 오픈함으로써 더 많은 사람이 이용할 수 있게 지원하고 있다.

이외에도 HSCIC는 데이터 이동 시 필수적인 보안통신 프로토콜 지원, 환자-의료진-사업자와의 연계, 전자 의료 기록 전송, 환자 의료 기록 관리, 전자 처방전 서비스 등 보건의료정보와 관련된 대부분을 담당하고 있다.

2) 미국

미국은 질병통제예방센터(CDC, Centers for Disease Control and Prevention)가 예방과 건강증진을 위한 콘텐츠를 생산해 직접 자국민에게 제공하며 추가로 사업자들이 콘텐츠를 고객에게 서비스할 수 있도록 오픈 API 서비스도 지원하고 있다. 기업이나 개인은 CDC 콘텐츠 라이브러리에서 원하는 콘텐츠를 API 형태로 제공받을 수 있다. 국내에도 유사한 형태로 보건복지부가 운영하는 국가건강정보포털 서비스가 존재하지만, 백과사전 형태의 콘텐츠만 등록되어 있어 최신 이슈와 시즌별 이슈에 대응할 수 있는 정기 업데이트 뉴스성 콘텐츠는 등록되지 않아 미국과 영국에 비해 제한적인 상황이다.

그림 3-1-13 CDC 콘텐츠 신디케이션 모델



※ 출처: <http://www.cdc.gov>

나. 민간 서비스 현황

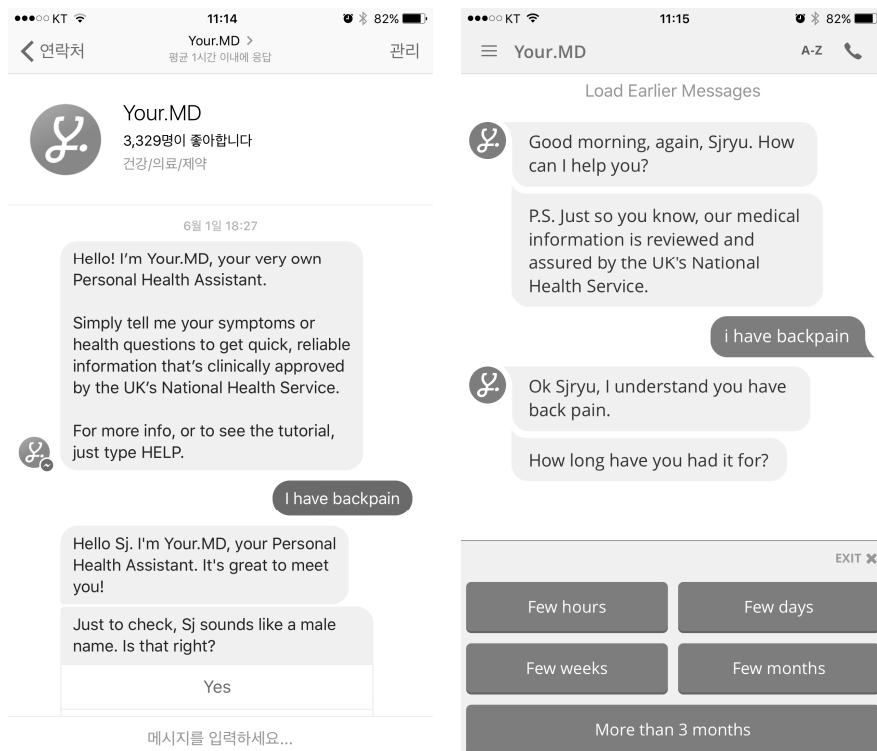
1) 영국

Your.MD는 최근 영국에서 가장 떠오르고 있는 보건의료정보 분야 신규 서비스이다.

알파고 붐으로 자국의 인공지능 기술을 알린 영국이 보유한 또 하나의 건강 관련 인공지능 솔루션인 Your.MD는 수많은 건강관련 DB가 축적된 인공지능이 사용자와 증상에 대한 대화를 하면서 스무고개를 풀듯 해당 사용자의 질병을 유추해 나간다. 질병이나 증상이 특정될 경우 NHS Choices로부터 제공받은 콘텐츠를 사용자에게 제공하고, 서드파티 사업자와 제휴를 통해 의료가 관과 연결해 주고 있다.

데이터의 집합체인 인공지능 중심 서비스로 자체적으로 운영하고 있는 애플리케이션 외에도 페이스북 메신저(Facebook messenger), 텔레그램(Telegram), 슬랙(Slack), 라인(Line(예정)) 등 외부 채팅서비스에 인공지능 DB API를 제공하며 사용자 확보와 인공지능의 정확성을 높여가고 있다.

그림 3-1-14 Your.MD 서비스 화면(페이스북 등)



///

최근 Your.MD는 인공지능 기반의 채팅 서비스와 함께 주변 병원 검색, 예약 등 보건의료정보를 통합하여 제공하기 위해 개발 중이며, 세계 모든 나라에서 해당 서비스를 제공할 수 있도록 우리나라를 포함한 여러 나라와 제휴를 진행하고 있다.

2) 미국

최근 오바마케어로 변화가 나타나고 있지만 미국은 전통적으로 보건의료 분야에서 민간의 영역이 압도적으로 높았다. 이에 따라 보험회사 등과 사업 제휴를 통해 콘텐츠 DB 중심의 보건의료정보 서비스를 제공하는 WebMD가 세계 최고의 보건의료정보 데이터 사업자로 자리매김하게 되었다.

WebMD는 자체 서비스 및 Medscape, MedicineNet, RxList 등 보건의료 전문분야의 콘텐츠 서비스를 운영하고 있다.

모바일 플랫폼에서는 Healthtap과 Zocdoc이 보건의료정보에서 압도적인 성과를 나타내며 시장 리더로 떠오르고 있다.

Healthtap은 의사와 사용자를 연결시켜 주는 모바일 헬스 커뮤니케이션(Health communication) 전문 서비스로 현재 10만명 이상의 미국 의사가 참여하고 있으며 지금까지 누적 47억건의 건강상담 서비스가 이뤄졌다. Healthtap은 이 성과와 확보된 DB를 바탕으로 외부 제휴를 넓혀가고 있다.

Zocdoc은 예약 시간이 비어 있는 의사와 지금 급하게 병원을 방문해야 하는 환자의 요구를 충족시켜 주는 의료 서비스 공급자 DB 중심 병원 예약 서비스로, 미국 전역의 병원과 의사 정보를 고객에게 제공하며 예약 중개를 통해 수익을 창출하고 있다.

4. 향후 전망

보건의료 분야는 헬스케어라는 세계적으로 가장 잠재력이 높은 시장 한가운데 위치해 있다. 모바일 플랫폼 시대에는 국가간 장벽도 없어 글로벌 사업이 가능한 분야이기도 하다. 실제로 아마존이 운영하는 글로벌 웹사이트 트래픽 분석 서비스인 알렉사(Alexa)에 따르면 webmd(세계 1위 건강의학콘텐츠 서비스)의 홈페이지 방문자 비율은 미국이 62%, 인도가 9.5%, 영국이 3.1%, 캐나다 3%, 중국이 2.1%로 전세계에서 동시에 이용되고 있다.



표 3-1-7 webmd의 국가별 점유율과 해당 국가 내 순위

	미국	인도	영국	캐나다	중국
점유율	62%	9.5%	3.1%	3%	2.1%
순 위	108	350	390	210	1,714

※ 출처: 알렉사(Alexa) 홈페이지

헬스케어는 질병을 예방하고 건강한 삶을 영위하는 웰니스 분야와 질병을 정확하고 빠르게 치료하는 메디컬 분야로 나눌 수 있다.

위 두 가지 분야를 모두 정확하게 수행하기 위해 보건의료정보는 필수적으로 요구된다. 웰니스 분야는 인구 통계학적 데이터, 생활 습관 유형, 센서를 통해 수집된 라이프로그 등의 데이터를 통합해 사용자의 현재 상황을 분석하고, 결과에 맞춰 맞춤형 콘텐츠와 전문가를 연결하는 데이터 서비스가 핵심이다. 과거에는 센싱 즉, 디바이스 개발이 필수적이었으나 최근에는 스마트폰을 포함한 대부분의 웨어러블 디바이스가 API를 오픈해 주고 있어 데이터를 관리하는 기술만 있더라도 충분히 웰니스 분야의 서비스를 만들어 낼 수 있다.

메디컬 분야는 진료가 필요한 상황에서 대상자에게 적합한 진료과와 병원을 추천해 주고, 진료 후 사후관리를 제공하는 서비스가 필요하다.

심평원을 비롯한 대부분의 공공기관은 지속적으로 보다 고도화된 정보를 제공하기 위해 노력하고 있다. 하이닥, 굿닥 등 전문 서비스가 네이버, 다음 못지 않게 많은 병원DB를 확보하게 된 것도 이 때문이다. 향후 우리나라에도 Webmd, Your.MD, Healthtap, Zocdoc과 같은 세계적인 서비스가 나올 수 있는 환경은 열려 있다.

제4절 신용정보

1. 신용정보 개요

가. 신용정보의 정의 및 분류

신용정보법(신용정보의 이용 및 보호에 관한 법률, 이하 신용정보법) 제2조에서는 신용정보를 ‘금융거래 등 상거래에 있어서 거래 상대방의 신용을 판단할 때 필요한 정보’로 정의하고 있다. 구체적으로는 식별 정보, 신용 거래 정보, 신용도 판단 정보, 신용 능력 정보, 공공 정보 등이 해당되며, 개인 신용정보는 생존하고 있는 개인의 정보로서, 사망자 정보는 개인 신용정보에 해당되지 않는 점에 유의할 필요가 있다.

신용정보는 개인 신용정보와 기업과 관련된 신용정보인 기업 신용정보, 기술에 관한 정보를 활용하여 새로이 만들어지는 정보인 기술 신용정보 등으로 분류되기도 한다.

신용정보는 전통적으로 부정적 정보와 긍정적 정보로 분류할 수 있는데, 부정적 정보(Negative Information)는 연체, 부도 등과 같이 신용정보 주체에 대한 신용평가에 불리한 영향을 미치는 정보, 긍정적 정보(Positive Information)는 대출 상환 실적, 소득 금액 등 신용평가에 유리한 영향을 미치는 정보라고 할 수 있다.

최근에는 Thinfile⁵⁴⁾고객⁵⁴⁾에 대한 신용평가에 SNS 정보 등이 활용되기도 하는데, 이러한 비전통적인 정보를 대체정보(Alternative Data)라고 하여, 최근 P2P 대출업체, 모바일뱅킹 등에서 적극 활용되고 있다.

나. 신용정보 활용 목적

신용정보를 개인 및 기업의 신용평가에 활용하는 이유는 경제 주체들로 하여금 개인 및 기업의 신용도에 대한 예측을 가능하게 하여, 정보의 비대칭성을 완화시킴으로써 경제활동의 사회적 비용을 감소시키는 데 있다.

금융시장에서 정보의 비대칭성은 금융회사로 하여금 역선택(Adverse Selection : 은행의 대출에 신용도가 낮은 고객이 보다 적극적)과 도덕적 해이(Moral Hazard : 대출 이후 상환을 하지 않을 유인 발생)를 야기하여, 선량한 금융소비자에게 피해를 주고 금융산업의 건전성을 저해하게 된다. 그러나 금융회사들이 신용정보를 적절하게 활용한다면, 금융소비자의 신용도를 정확하게 판단할 수 있게 되고 이에 따라 대출 여부와 수준(금액과 이자율 등)을 적절하게 결정할 수 있게 된다. 이는 금융산업뿐만 아니라 신용이 발생하는 일반 산업에서도 동일하게 적용될 수 있다.

⁵⁴⁾ 신용정보가 부족하여 신용평가등급이 산출되지 않는 고객군을 통칭하는 말로, 대학생, 주부 등이 여기에 해당된다.

다. 신용정보 공유 체계

신용정보법상의 신용정보의 공유 체계 참여자는 [표 3-1-8]과 같다.

표 3-1-8 신용정보법상 신용정보 공유 체계

구 분		내 용
신용정보 집중기관	종합신용정보 집중기관	• 대통령령으로 정하는 금융기관 전체로부터의 신용정보를 집중관리·활용 • 한국신용정보원
	개별신용정보 집중기관	• 상기 금융기관 외 같은 종류의 사업자가 설립한 협회 등의 협약에 따라 신용정보를 집중관리·활용 • 2016년 이전에는 한국여신금융협회, 생명보험협회, 한국금융투자협회 등이 해당되었으나, 2016년부터는 한국신용정보원으로 정보가 통합 집중되어, 상기 기관들은 단순 협회로서의 기능만 가지고 있음
신용정보회사	신용조회업무	• 개인, 기업에 대한 신용정보를 수집하여 자체 신용평가모형을 통해 부도 위험 등을 나타내는 신용평점·신용등급을 산출하고, 신용평점 및 기타 신용정보를 대출거래 등에 필요시 제공 • 코리아크레딧뷰로, NICE평가정보, SCI평가정보, NICE디앤비, 이크레더블, 한국기업데이터 등
	신용조사업무	• 타인의 의뢰에 따라 신용정보를 조사하여 제공 • 한국TDB신용정보
	채권추심업무	• 채권자의 위임을 받아 연체한 채무자에 대하여 재산조사, 채무변제촉구, 변제금 수령 등 추심채권을 행사 • SGI신용정보, 고려신용정보, 미래신용정보, 신한신용정보 등
	신용정보 주체	• 신용정보로 식별되는 자로서, 그 신용정보의 주체가 되는 자
신용정보제공·이용자		• 고객과의 금융거래 등 상거래를 위하여 본인의 영업과 관련하여 얻거나 만들어낸 신용정보를 타인에게 제공하거나 타인으로부터 신용정보를 제공받아 본인의 영업에 이용하는 자로서 주로 금융기관 등이 포함

2. 개인 신용정보 서비스

가. 금융회사 대상 서비스

기업(주로 금융회사)에 제공하는 신용정보회사의 개인 신용정보 서비스는 크게 세 가지 영역으로 요약할 수 있다. 이 중에서 신용리스크 관리 서비스는 이미 전 금융회사에서 대출, 카드 등의 심사업무에 활성화 되어 사용되고 있으며, 최근에는 부정위험관리, 전략적 의사결정 지원시스템 등의 구축도 활발해지고 있다.

개인 신용정보 서비스에 활용되고 있는 신용정보는 식별정보, 신용거래정보, 신용도 판단정보, 신용능력정보 및 공공정보 등이 있다. 특히, 공공정보 중 과태료 및 세금체납 등의 정보는 신용리

스크관리 영역에서, 관보정보(법원의 금지산, 한정재산, 실종 선고, 회생 및 파산 등)는 부정거래 위험관리 서비스 영역에서 활용되고 있다.

표 3-1-9 금융회사 대상 개인 신용정보 서비스 영역

구 분	내 용
신용리스크관리	• 금융회사가 개인/개인 사업자에게 대출을 시행함에 있어 발생할 수 있는 채무불이행 위험을 정확히 측정할 수 있도록 제공하는 서비스로서, 개인의 신용거래 내역 등을 분석하여 불량률을 예측하는 개인신용평점(Score) 서비스가 대표적
부정거래 위험관리	• 금융회사가 신용사기 거래를 개설 시점이나, 개설 후 시점에서 예방하거나, 적발할 수 있도록 지원하는 서비스
전략적 의사결정지원	• 전체 가계신용 시장에서의 자사의 위치를 파악하고, 전반적인 경제 트렌드 속에서 자사의 현재 전략을 점검할 수 있도록 지원하는 서비스로서, 금융회사는 이 서비스를 통해서 시장환경에 맞는 전략을 수립

나. 개인 대상 서비스

신용정보회사는 온라인을 통한 개인의 신용정보를 직접 확인하고 관리할 수 있는 신용관리 서비스를 운영하고 있다. 개인에게 제공되는 주요 정보로는 신상정보, 대출·카드정보, 보증정보, 연체정보, 조회정보 등이 있으며, 이외에도 신용위험도 진단, 신용대출 진단, 금융명의 보호, 신용수준 비교 등의 서비스를 제공하고 있다.

본인이 자신의 신용정보를 조회할 경우에는 신용도 하락이 전혀 없으며, 이를 통해 신용평점, 신용등급, 주요 변동내역 등 본인의 신용을 상세히 파악하고 관리할 수 있다. 또한 각 서비스 사이트는 1년 3회에 한해 전국민이 무료로 신용정보를 조회할 수 있도록 무료조회 서비스도 제공하고 있다.

특히, 신용등급 및 신용정보가 개인의 경제생활에 미치는 영향도가 점차 확대됨에 따라, 단순히 본인의 신용을 확인하는 데서 그치지 않고 본인의 신용을 적극적으로 관리하고 상승시키려는 요구가 커지고 있다. 이에 따라 신용정보 회사에서도 신용정보 변경에 따른 신용등급을 예측할 수 있는 ‘신용등급 시뮬레이터’, 전문가가 직접 신용등급 향상 방법을 제공하는 ‘신용등급 향상 1:1 코칭 서비스’ 등을 제공하고 있다.

표 3-1-10 개인 대상 신용정보 조회서비스 운영 사이트

구 분	사이트 주소(URL)	회사명
올크레딧	www.allcredit.co.kr	코리아크레딧뷰로(KCB)
NICE지키미	www.credit.co.kr	NICE평가정보
사이렌24	www.siren24.com	SCI평가정보

3. 기업 신용정보 서비스

기업신용정보 서비스는 기업의 정보를 확인하는 기업정보조회 서비스와 기업의 금융상품 및 신용공여 등(회사채, 기업어음, 유동화증권)의 원리금이 상환될 가능성 등 기업 및 법인, 펀드 등의 신용도를 평가하여 평가결과를 제공하는 신용평가 서비스로 분류된다.

표 3-1-11 기업 신용정보 서비스 영역

구 분	내 용
기업정보조회 서비스	• 기업의 재무제표, 재무현황, 기업현황, 주주현황, 관계사 현황 등의 기업 상세 정보를 제공하는 서비스
기업신용평가 서비스	• 회사채 등 유가증권의 발행, 거래시의 투자판단 및 발행조건 결정의 기준 및 거래기업의 적격 심사 등 신용판단을 위한 기준으로 활용되는 기업에 대한 신용등급을 제공하는 서비스

4. 향후 전망

신용정보법이 2016년 3월 12일 개정 시행됨에 따라서, 신용조회회사(민간 CB사)의 업무 범위가 대폭 축소된 반면, 신용정보에 대한 관리, 보안이 대폭 강화되고, 그동안 개별 신용정보집중기관(개별 협회)에서 수집, 활용되던 카드/보험/증권 정보가 한국신용정보원으로 통합 집중 관리되게 되었다.

최근에는 ‘개인정보 비식별 조치 가이드라인’⁵⁵⁾이 마련되어, 익명화된 신용정보를 핀테크업체, 금융회사들이 자유롭게 활용할 수 있는 길이 열리게 되었다.

이에 향후 신용정보는 기존 리스크관리 분야뿐만 아니라, 다양한 업권에서 다양한 목적으로 활용될 수 있을 것으로 기대된다. 하지만 신용조회회사(민간CB사)의 업무 역할이 본인식별, 신용평가/위험관리 모형개발로 국한되고 공적기관인 한국신용정보원의 업무 역할이 확대된 점을 고려한다면, 신용정보의 정교한 활용 측면에서는 다소 아쉬운 점이 있다는 업계의 분석이다.

55) 2016년 6월 말 국무조정실, 미래창조과학부, 행정자치부, 방송통신위원회, 금융위원회, 보건복지부 등이 공동으로 마련한 가이드라인으로서, 빅데이터 활용에 필요한 개인정보의 비식별 조치 기준과 절차, 방법 등이 정의되어 있다.

제5절 특허정보

1. 특허정보 개요

특허는 발명자가 연구개발한 새로운 기술사상을 공개하면 그에 대한 대가로 독점적 실시를 특별히 인정해 준다는 취지에서 출발한 제도이다. 특허가 포함된 지식재산권은 사람의 지식으로부터 파생된 창작물에 대한 권리와 배타적 소유권으로 단순히 기술을 보호할 뿐만 아니라 고부가가치를 창출하는 원천이 될 수 있어 그 중요성이 점차 높아지고 있다. 특히 최근 상표·디자인에 대한 글로벌적 소송과 분쟁이 발생함에 따라 특허뿐만 아니라 지식재산권 전체에 대한 관심과 활용이 높아졌고, 매년 그 출원 건수도 증가하고 있다.

더욱이 최근 지식재산권의 확보가 기업 및 국가의 중요한 경쟁력으로 작용하면서 심판정보, 소송정보, 권리변동 정보 등의 부가정보 제공의 필요성이 커지고 있다. 지식재산권 정보를 활용하여 새로운 기술과 제품을 개발하고 권리를 획득함으로써 기술개발을 독점하여 기업의 경쟁력을 높일 수 있다. 또한 기술정보·경영정보 또는 권리정보의 기능을 가지고 있으며 기업의 전략 수립과 관련하여 그 정보적 가치가 크다고 할 수 있다. 해외 선진 기업들과의 기술격차 해소 및 기술이전을 효과적으로 수행하기 위해서는 지식재산권 정보를 효율적으로 활용하는 것이 중요하다.

표 3-1-12 지식재산권 종류

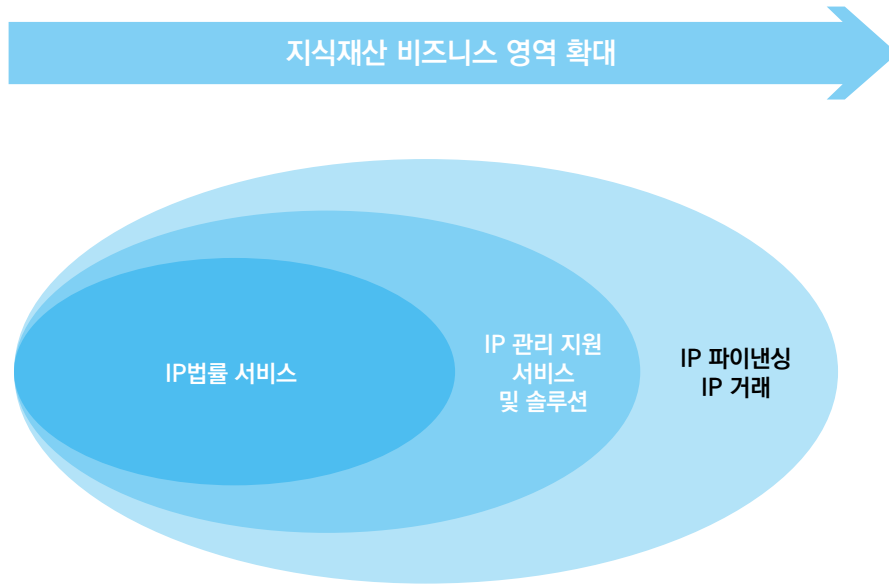
지식재산권 종류		개요
산업재산권	특허권	제조방법, 생산방법 등 높은 수준의 기술발전을 보호하는 권리
	실용신안권	작은 발명이나 아이디어 상품을 보호하는 권리
	상표권	기호나 문자, 색채 등에 의해 상품을 표시하는 것에 대한 브랜드명을 보호하는 권리
	디자인권	어떠한 물품에 대한 외형적 디자인을 보호하는 권리
저작권	저작권	문화·예술 분야의 창작물에 대해 부여하는 배타적인 권리
	저작인접권	

이러한 특허정보를 보다 쉽게 검색하고 효율적으로 활용할 수 있도록 특허정보를 데이터로 구축한 것이 특허정보 데이터베이스이다. 특허정보는 원천 데이터가 비정형화 되어 있고, 국가별/시기별/대상별로 다양한 형태로 배포되는 것이 특징이지만 전세계적으로 통일된 표준형식을 규정하고 있어 비교적 동일한 형태의 글로벌적 데이터베이스 구축과 활용이 가능하다.

또한 지식재산 비즈니스의 영역이 지식재산 등록, 라이선싱 계약 및 소송에서 IP금융, IP법률 서비스, 가치평가 등으로 확대가 되면서 특허정보 검색 중심의 서비스에서 분석·평가·무효화 등

다양한 형태의 서비스가 제공되고 있다.

그림 3-1-15 지식재산 비즈니스 영역



※ 출처: 지식재산서비스산업의 국내외 현황 연구, 특허청, 한국지식재산서비스협회, 2012

2. 국내 서비스 현황

가. 공공 서비스 현황

지식재산 비즈니스 영역이 확대되면서 공공의 특허정보 서비스 유형도 다양해지고 있다. 과거 온라인 정보 제공이 주요 서비스 형태였다면, 정부 3.0 도입으로 공공에서의 데이터 보급 서비스가 더욱 적극적이고 활발해지고 있다. 뿐만 아니라 공공기관 간 정보 공유를 통해 주제가 다른 콘텐츠에 대해 융합형 정보를 제공함으로써 데이터 활용성을 높이고 있다. 특히 특허청의 데이터 보급은 국내 데이터에 한정되어 있는 것이 아니라 해외 특허청의 데이터를 적극적으로 보급하여 국내 기업들의 글로벌 경쟁력을 확보할 수 있는 기회를 제공하고 있다.

1) KIPRIS Plus(<http://plus.kipris.or.kr>)

특허청이 개방 중인 모든 산업재산권 정보를 데이터셋(Bulk Data), 오픈 API, LOD 방식으로 실시간 제공하는 서비스이다. 초기 공공데이터 중심에서 민간기업의 데이터까지 보급 대상을 확대하고 있다. 2015년 기준 공공데이터 137개, 민간데이터 47개가 등록되어 약 3억 9,000만건 이상의 데이터를 제공하고 있다. 'KIPRIS Plus' 서비스의 제공으로 지식재산 관련 업계뿐만 아니라 대학, 연구기관, 금융권, IT 중소기업 등에서 IP서비스, IP금융, IP거래, IP관리, IP 평가 분야에 다양하게 지식재산 분야의 정보를 활용할 수 있게 되었다.

그림 3-1-16 KIPRIS Plus 특허정보 활용 서비스

The screenshot shows the KIPRIS Plus website interface. At the top, there's a navigation bar with 'DATA', 'OPEN API', 'LOD', '개발자광장', '참여소통', 'KIPRIS Plus', and 'HelpDesk'. Below this is a large banner for 'KIPRIS Plus Help Desk Open' with a sub-header '고객님의 소중한 인권을 등록해주세요.' and a '바로가기 >>' button. To the right of the banner is a 'NEWS & NOTICE' section with a list of recent updates and dates.

The main content area is divided into three columns. The left column is titled 'DATA' and contains sections for 'DATA 개방현황' (Data Opening Status), '출처별 개방 현황' (Opening Status by Source), and '데이터 상품 상세 현황' (Detailed Status of Data Products). The middle column is titled '최근 등록 데이터' (Recently Registered Data) and lists various data categories like '국가연구개발사업정보', '중국 등록공보', '중국 공개공보', '주요국 가전 등록디자인', '무료 상품 정보', '남녀별 등록건수 통계 정보', '아이디어DB', '국가별 등록건수', and '공익표창'. The right column is titled '기업 DATA 정보' (Company Data Information) and lists various services like '컴파스(Compass)', 'i-IPS(특허관리시스템)', '마크서치 모바일', '특허법률사무소관리(IPAMS)', and '(주)비전인사이드의 R&D 특...'. The bottom of the page has a footer with the text '※ 출처: <http://plus.kipris.or.kr>'.

2) IP-NAVI(<http://ip-navi.or.kr>)

한국지식재산보호원에서 운영하고 있는 ‘IP-NAVI’는 국제 지재권분쟁정보 포털사이트로 주요 국가의 지재권 관련 판례 및 특허, 상표, 디자인 등 산업재산권 관련 분쟁 정보를 제공하며 독창적인 기술 분류체계와 검색 서비스를 제공하고 있다. 최근 온라인검색 서비스뿐만 아니라 국내의 판례정보의 데이터 보급까지 서비스 대상을 확대하고 있다.

그림 3-1-17 IP-NAVI 국제 지재권 분쟁 정보 포털



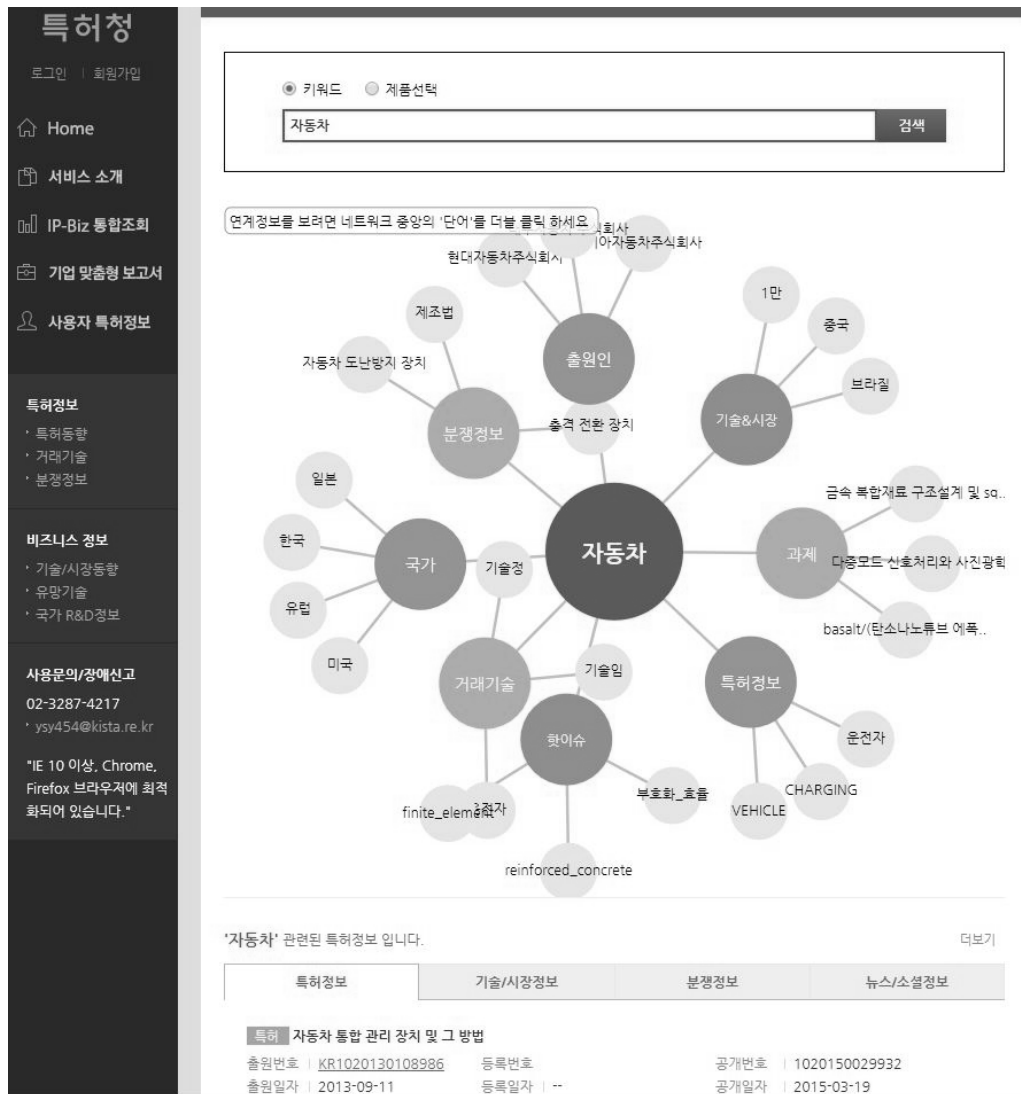
※ 출처: <http://www.ip-navi.or.kr>

3) IP-BIZ정보 하나로 서비스(<http://ip-biz.kr>)

특허청이 보유한 지식재산권 정보를 중심으로 타 정부부처 및 공공기관이 보유 중인 다양한 비즈니스 정보를 연계하여 활용할 수 있는 정보채널을 제공하는 서비스로 IP·기술·시장·R&D 영역을 포괄하는 통합 정보 서비스이다. 과학기술정보, 기업정보, 거래기술 정보 등 약 21개의 다양한 주제의 콘텐츠를 연계하고 있다. 주요 제공 서비스로는 기업맞춤형 보고서, 특허동향, 거래기

술, 분쟁정보, 기술/시장동향 정보 등이 있다. 특히 최근 IT 기술을 반영하여 정보 접근성을 용이하게 하고 있다.

그림 3-1-18 IP-BIZ정보 하나로 서비스



※ 출처: <http://ip.biz.kr>

나. 민간 서비스 현황

국내 특허정보 비즈니스 시장은 해외 시장에 비해 그 규모가 크지 않아 기존 서비스 업체들을 중심으로 특허정보 온라인검색 서비스, 연차료 관리 서비스, 상표 서비스 등 특정 주제에 대한 서

비스를 제공하고 있던 서비스 모형이 점차 포털 서비스 형태로 변화하고 있다. 또한 해외 서비스의 국내 진출이 활발해지면서 글로벌 시장과의 경쟁력 확보를 위해 콘텐츠 확대와 서비스 개선 등 시장 변화에 대응하기 위한 서비스 변화가 활발하게 이루어지고 있다.

1) WIPS ON(<http://www.wipson.com>)

전세계 특허정보 온라인 검색 서비스를 개발한 (주)웍스의 WIPS ON은 기존 특허문헌 중심의 콘텐츠에서 주요국의 상태 정보, 현재 권리자, 존속기간(예상) 만료일, 양수양도이력, 연차료 정보 등의 고급 행정정보와 심판정보, 소송정보를 제공함으로써 다양한 분야의 비즈니스에서 활용할 수 있는 서비스를 제공하고 있다. 또한 다양해진 콘텐츠를 기반으로 특허검색을 넘어 활용에 초점을 맞추어 아이디어 발굴, 무효자료조사, 회피 설계 등 목적별 최적화 서비스를 제공하고 있다.

그림 3-1-19 WIPS ON 홈페이지 화면

WIPS ON

기본검색 번호검색 스텝검색 통합검색 Beta 추천검색 전문가검색

아이디/비밀번호 찾기 회원가입

로그인

아이디/비밀번호 찾기 회원가입

쉽고 정확한 전세계 특허 검색서비스

국내 최대의 신뢰도 높은 전세계 특허 데이터베이스

공지사항 more +

- NEW 7/20(수) 00:00 ~ 07:00 서비스 일시 중지...
- 6/22(수) 00:00 ~ 06:00 서비스 일시 중지 안내
- 개인정보취급방침 변경 안내(Ver.1.8)
- KR 공개문헌 검색 장애
- NEW WIPS ON 정식서비스 오픈 안내

서비스개선 more +

- WIPS ON vs NEW WIPS ON 검색 결과 비교 안내
- NEW WIPS ON 신규기능 안내

IP 교육안내 more +

- [2월 : 2016.07.21~]R&D를 위한 특허정보검색...
- [1월 : 2016.07.08~]R&D를 위한 특허맵 작성 실무
- [2월 : 2016.04.21~]R&D를 위한 특허정보분석...
- [1월 : 2016.02.26~][유망신기술세미나]스마트...

웍스온 활용 TIP

다국가에 출원된 명세서 특허 한번에 검색하기

- 통상검색
- 중핵검색
- 패밀리 그룹링 활용방법

포대서비스 신청
특허등록 전과정 및 법률사항 정보

Blog
웍스의 아이디어 놀이터

WIPS 토탈 서비스 안내
지식재산 정보를 한곳에

서비스소개
이용안내

기능별 매뉴얼
FAQ

고객센터 02-726-1100 / 1105

팩스 0303-0955-1105 | 이메일 help@wips.co.kr
평일 09:00~12:00, 13:00~18:00 토요일 공휴일은 휴무입니다.

회사소개 | 이용약관 | 개인정보취급방침 | 고객센터

Family Site

(주)웍스 121-904 서울시 마포구 성암로 189 중소기업DMC타워 18층, 19층 | 사업자등록번호 : 105-86-02802 | 대표 : 이철철
Copyright © 2016 WIPS Co.Ltd. All rights reserved.

※ 출처: <http://www.wipson.com>

2) ANYFIVE(<http://www.anyfive.com>)

연구개발(R&D) 및 지식재산 서비스 전문기업인 애니파이브는 IPR(Intellectual Property-R&D) 및 ICT 솔루션과 기술을 제공하는 기업으로 컨설팅, R&D 통합관리, 지식재산권관리, 기술사업화 및 포트폴리오 서비스, 빅데이터 서비스 등을 제공하고 있다. 최근 해외 특허정보 검색 서비스와 제휴를 통해 100여 개국 특허를 쉽고 빠르게 검색할 수 있는 서비스를 국내시장에 도입하였다.

3) 마크프로(<http://www.markpro.com>)

국내외 특허권의 연차료 납부대행 업무를 목적으로 설립된 마크프로는 이후 기술이전 및 사업화, 기술정보조사, 국내외 상표검색 서비스 등 토털 IP관리 서비스를 제공하고 있다. 특히 2016년에는 국내지식재산 판례정보를 서비스하는 IPRs Case를 런칭하였다. 특허심판원뿐만 아니라 특허법원, 대법원의 판례정보를 쟁점별로 분리하여 서비스함으로써 최근 이슈가 되고 있는 특허소송 등에 유익한 정보를 제공하고 있다.

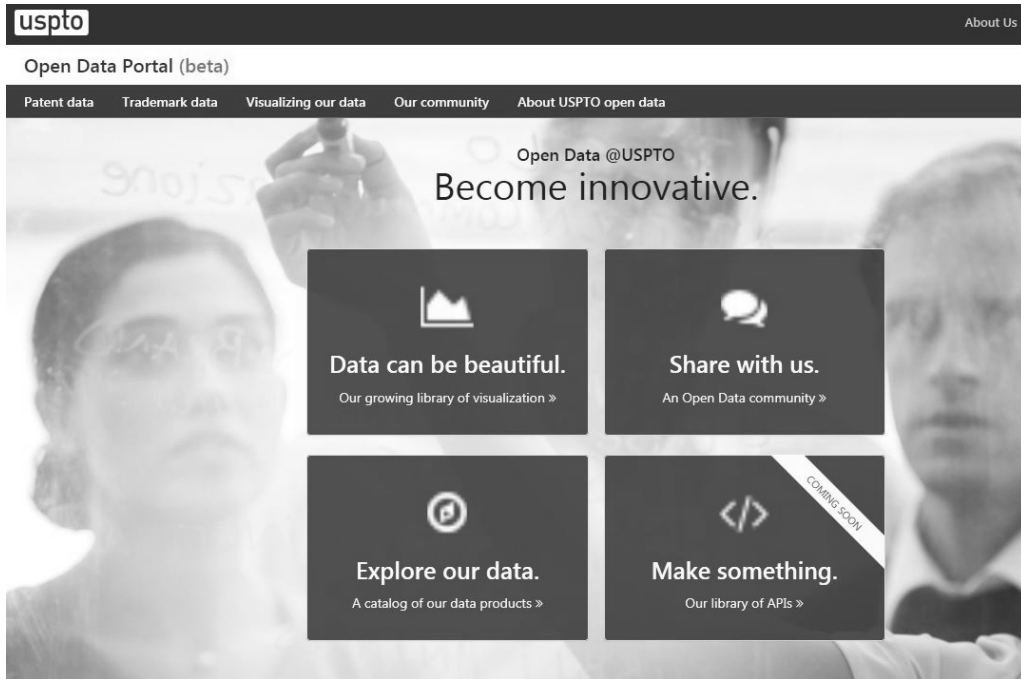
3. 해외 서비스 현황

가. 공공 서비스 현황

1) 미국 특허청

IP5(미국, 일본, 유럽, 중국, 한국) 특허청 중에서 데이터 보급이 활발하고, 유료 데이터에 대한 무료 보급 전환이 빨랐던 미국 특허청은 지속적으로 온라인을 통한 정보제공 서비스를 확대해 나가고 있으며, 미국 지식재산 데이터의 보급을 위한 다양한 형태의 서비스도 개발하고 있다. 특히 미국 특허청은 지식재산권 관련 소송 및 판례 정보 등 많은 이슈가 되고 있는 특허소송에 대응할 수 있는 정보를 타 특허청에 비하여 적극적으로 제공하고 있다.

그림 3-1-20 USPTO 오픈 데이터 포털 - 데이터보급 서비스

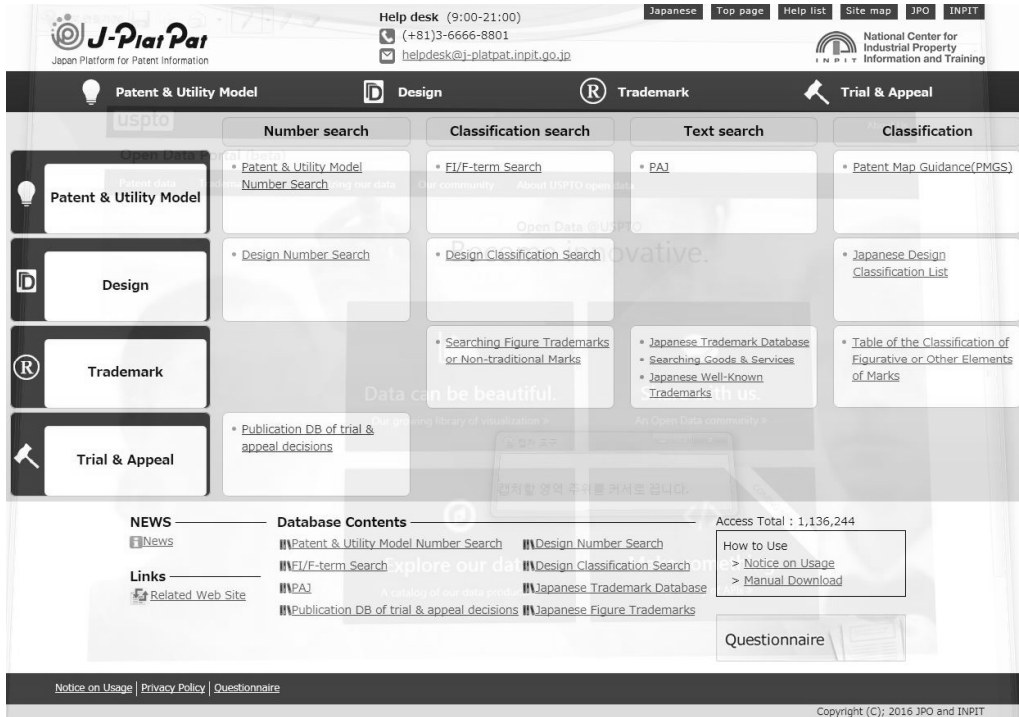


※ 출처: [http:// www.uspto.gov](http://www.uspto.gov)

2) 일본 특허청

과거 자국의 지식재산정보 중심 서비스를 제공했던 일본은 지식재산권이 글로벌화되면서 해외에서의 지식재산권 활동을 지원하고 자국민을 보호하기 위해 해외 지식재산권 정보 서비스를 강화하고 있다. 특히 2015년 이후 국내외 특허정보 데이터 보급이 활발해지고 있으며 무료로 데이터를 보급하고 있다. 다른 특허청과의 큰 차이점은 자국민을 우선으로 데이터가 보급되고 있다는 점으로, 해외 고객의 경우 제한적이거나 기존에 특허정보를 구매한 이력이 있는 경우만 데이터를 보급하고 있다. 또한 언어 장벽을 해결하기 위해 해외특허정보에 대한 일본어 번역 서비스 대상 국가를 확대하고 있다.

그림 3-1-21 J-PlatPat 지재권정보 검색 서비스



※ 출처: <http://www.j-platpat.inpit.go.jp>

4. 향후 전망

최근 특허정보 서비스 동향을 보면 정보의 글로벌화와 해외 기업 간의 특허소송이 많아지면서 정보와 서비스 모델이 더욱 다양하고 고급화되고 있다. 과거에는 주로 과거의 특허정보 조회나 검색 위주의 정적인 데이터가 서비스되었다면, 최근에는 소송, 심판, 판례, 권리자 변동정보, 존속기간(예상) 만료일 등의 동적인 데이터들을 활용한 서비스가 출시되고 있다.

한편, 특허정보는 그 기반이 공공데이터이며 데이터의 구축과 운영에 많은 비용이 투자되므로 기본적으로 공공의 데이터 보급 정책이 관련 서비스산업에 큰 영향을 미치게 된다. 우리 정부는 정부 3.0 도입 이후 많은 데이터를 민간에게 개방하였고, 앞으로도 시장의 수요에 따라 지속적으로 다양한 주제의 정보 보급이 예상된다.

민간시장에서는 이러한 정부의 정책변화에 따른 다양한 주제의 데이터들을 융합하여 새로운 가치의 정보 서비스 모델을 발굴할 것이며, 이로써 새로운 정보 이용자층의 확대는 물론 기존 이용자들의 정보 활용도가 더욱 높아지게 되어 그들의 비즈니스에 미치는 영향력도 커질 것으로 예상된다.

제6절 교육·훈련정보

1. 교육·훈련정보 개요

가. 서비스 정의

1990년대 이후 세계경제는 정보통신기술을 비롯한 과학기술 발달에 힘입어 빠른 속도로 지식기반경제(Knowledge-based Economy)로 전환되고 있으며 지식기반 서비스업의 중요성도 커지고 있다. 지식기반 서비스업이란 지식이 체화된 중간재를 생산 활동에 집적하여 고부가가치를 창출하는 산업으로 제조업에 비해 부가가치 유발계수가 높고 노동집약적 특성을 가지고 있어 고용창출 효과가 크다는 특징이 있다. 정보통신을 비롯하여 금융/보험, 의료, 문화, 교육 등이 지식기반 서비스업에 포함되며 최근에는 산업 간 융합분야인 에듀테크, 핀테크, 바이오테크가 유망 산업으로 떠오르고 있다. 교육 서비스는 국가인적자원 개발에 핵심적 역할을 하는 대표적인 지식기반 서비스로써 학교 교육과 평생 교육뿐 아니라 직업훈련, 국제교육 부문까지 범위가 매우 넓다. ICT 발전과 함께 전통적 집체훈련 방식은 이러닝, 블랜디드 러닝 등 다양한 형태로 진화하고 있다.

나. 서비스 구분

1) 표준산업분류에 따른 구분

표 3-1-13 교육 서비스 산업 표준분류

구 분	산업분류	교육제공기관	KSIC
학교 교육	초등 교육기관	유아교육기관 / 초등학교	P851
	중등 교육기관	중학교 / 일반 고등학교 / 기술 및 직업 중등 교육기관 / 상업 및 정보 산업 고등학교 / 공업 고등학교 / 기타 기술 및 직업 고등학교	P852
	고등 교육기관	전문대학 / 대학교 / 대학원	P853
	특수학교, 외국인학교 및 대안학교	특수학교 / 외국인학교 / 대안학교	P854
시설 교육	일반 교습 학원	일반 교과 학원 / 외국어학원 / 방문교육 학원 / 온라인 교육학원 / 기타 일반 교습 학원	P855
	기타 교육기관	스포츠 및 레크레이션 교육기관 / 예술 학원 / 사회교육시설 / 직원 훈련기관 / 기술 및 직업훈련학원 / 그외 기타 교육기관	P856
	교육지원 서비스업	교육지원 서비스업 / 교육관련 자문 및 평가업 / 기타 교육지원 서비스업(이러닝 포함)	P857

※ 출처: 제9차 한국표준산업분류 분류항목표(통계청고시 2007-53호), 통계청

통계청의 한국표준산업분류(KSIC, Korea Standard Industrial Classification)에 따르면 교육 서비스 산업은 크게 학교 교육과 사설 교육으로 나뉜다. 학교 교육은 초등교육, 중등교육, 고등교육과 특수학교 외국인학교 및 대안학교로 분류되며, 사설 교육은 일반 교습학원, 기타교육학원, 교육지원 서비스업으로 분류된다.

2) 국제표준교육분류에 따른 구분

국제표준교육분류(ISCED, International Standard Classification of Education)는 교육 프로그램을 구성하고 교육 수준과 분야에 맞는 관련 자격 요건을 위한 참조 분류체계이다. 2011년 유네스코(UNESCO)가 국제표준교육분류를 개정한 후 우리나라도 표준분류 제정에 착수하여 2015년부터 국제표준교육분류를 제공하고 있다.

한국표준교육분류는 교육 수준과 교육 영역을 기준으로 교육을 구분하고 있다. 교육 수준은 교육프로그램 분류와 교육이수 분류로 구분하고 각 단위를 대-중-소로 구성하였다. 대분류는 교육 수준을 기준으로 영유아 과정과 초등 ~ 대학원(석사, 박사 과정)까지의 교육과정에 따라 단계별로 분류하며, 중분류는 교육 성격을 기준으로 2단계 전기중등 과정(중학교) ~ 5단계 전문(산업)학사(또는 상당) 과정까지는 일반교육과 직업교육으로 구분하고, 6단계 학사(또는 상당) 과정 ~ 8단계 박사(또는 상당) 과정까지는 학문교육, 전문교육, 불특정 항목으로 분류하였다. 소분류의 경우, 교육프로그램 분류에서는 각 단계별 교육과정 완성 여부 및 상위 교육 단계로의 진학 자격을 기준으로 분류를 세분화하고, 교육이수 분류에서는 개인이 학습한 최상 수준의 교육 완성 여부를 기준으로 세분화하였다.

표 3-1-14 교육이수 분류체계

대분류 - 단계		중분류 - 성격		소분류 - 이수	
0	초등 과정 이전	0	별도 정의 없음	0	별도 정의 없음
1	초등 과정	1	교육 프로그램에 참여한 적 없음	1	재학
2	전기중등 과정(중학교)	2	영유아	2	중퇴
3	후기중등 과정(고등학교)	3	초등교육	3	휴학
4	중등 후 비학위 과정	4	일반/학문	4	수료
5	전문(산업)학사(또는 상당)과정	5	직업/전문	5	졸업
6	학사(또는 상당)과정	6	불특정		
7	석사(또는 상당)과정				
8	박사(또는 상당)과정				
9	기타	9	미상	9	미상

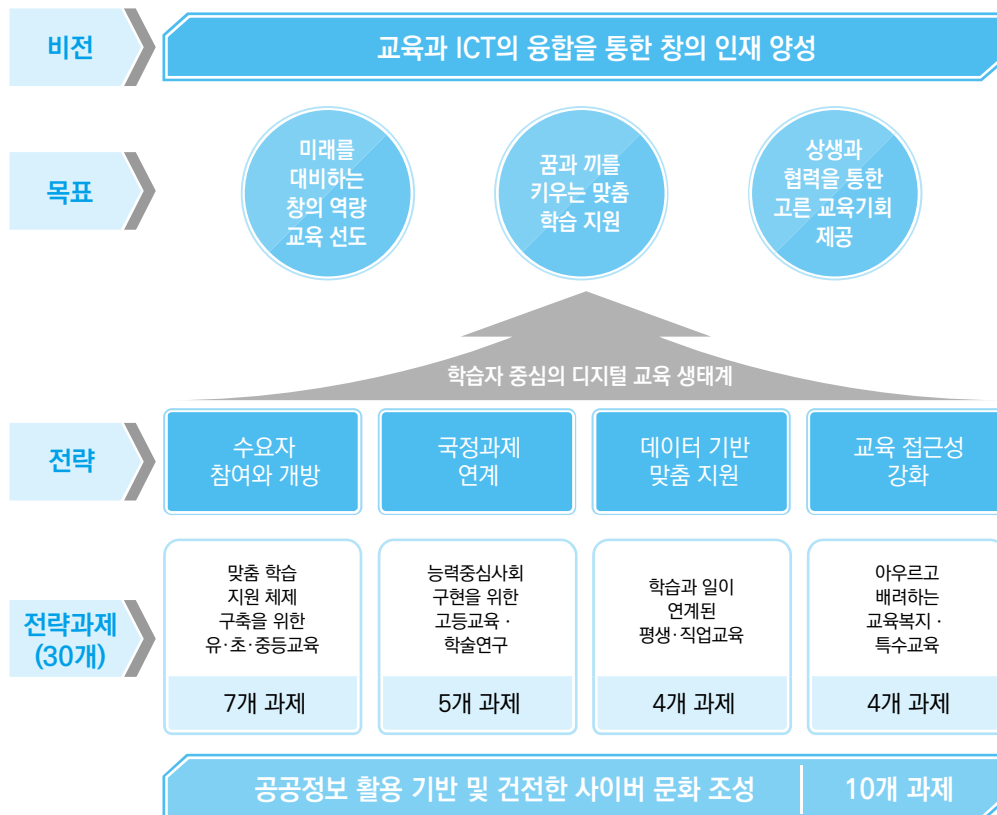
2. 공공 서비스 현황

국내 교육훈련 공공 서비스는 크게 학교 교육과 평생학습, 직업훈련으로 구분될 수 있으며 그 외 학술정보 서비스가 공공 서비스에서 큰 영역을 차지한다.

가. 학교 교육

교육부는 교육정보화 기본계획에 의거해 교육 관련 정책을 집행하고 있다. 제4차 교육정보화 기본계획이 완료됨에 따라 제5차 기본계획을 수립하여(2014~2018) ▶맞춤학습 지원체제 구축을 위한 유·초·중등교육 ▶능력중심사회 구현을 위한 고등교육·학술연구 ▶학습과 일이 연계된 평생·직업교육 ▶아우르고 배려하는 교육복지·특수교육 ▶공공정보 활용 기반 및 건전한 사이버 문화 조성 등 5대 전략과제와 그에 따른 30개 전략과제를 추진하고 있다.

그림 3-1-22 제5차 교육정보화기본계획(2014~2018) 개요



※ 출처: <http://library.krivnet.re.kr/search/detail/CATTOT000000056567>

교육 공공정보의 적극적인 공개로 국민의 알권리를 충족하고 공공데이터 민간 활성화를 촉진하기 위해 공공데이터 활용에 적극 나서고 있다. 공공데이터 개방 제도란, 정부 및 공공기관이 보유한 공공데이터를 민간에서 활용할 수 있도록 개방하는 제도로 교육부는 홈페이지를 통해 교육관련 공공데이터를 개방하고 있다.

표 3-1-15 교육부 제공 공공데이터 목록

공공데이터	내용
평생학습 개인실대조사	우리나라 성인의 평생학습 참여 현황
교과용 도서 수정·보완 현황	현행 교과서 관련 수정, 보완대조표, 수정 예정 사항, 교과서 모니터링 정보 제공
학생건강체력 평가제	초중고 학생의 학생 건강 체력 평가 정보
교과서 보드 DB	국정검정 교과서 정보 서비스 관련 과정 고시, 공지사항, 교과서연구, 성취기준 평가기준 등 cutis에 탑재되어 있던 내용 모두 게재
교과서 컴멘터리 DB	1차 이전부터 2009 개정과정까지 고시 및 과정 해설서
학교건강검사 표본조사결과	전국 초·중·고등학생의 신체발달, 건강검진, 건강조사 표본조사 결과
통일, 안보자료 서비스 조회수	통일, 안보 관련 수업자료 및 계기 자료를 제공하는 서비스
기본통계(대학교)	전국 대학교 현황(학교, 학교, 시설 현황)과 관련한 통계자료
특수통계	특수기관, 특수대상자, 특수교원 등 관련 인력 등 특수 관련 시도청별 현황
기본통계(유치원)	전국 유치원 현황(학교, 학교, 시설 현황)과 관련한 통계자료
교육기본통계(초,중,고등학교)	전국 초등학교, 중학교, 고등학교 현황(학교, 학교, 시설 현황)과 관련한 통계자료
유초중고등교육기관 기본현황	유·초·중고등교육기관 학생수, 교원수, 주소 등 현황
고등교육기관 졸업자 취업통계	한국교육개발원이 보유하고 있는 국내 전문대학, 대학, 교육대학, 산업대학, 각종학교, 기능대학, 일반대학원 각 학제별 취업현황(취업률, 진학률 등) 정보
평생교육통계	평생교육기관 현황, 운영 프로그램 현황, 교강사 및 사무직원 현황 등
대학특성화정보	대학특성화 종합정보
방과후학교 운영 현황	초·중·고 방과후학교 및 초등돌봄교실 운영현황 자료 제공
교과서 텍스트북 DB	교과서목록표

※ 출처: <http://www.moe.go.kr>, 교육부

교육통계정보를 제공하는 대표 기관으로 교육통계연구센터(<http://kess.kedi.re.kr>)가 있다. 교육통계연구센터는 국가 교육통계사업 수행을 전담하는 기관으로 한국교육개발원 소속 조직이다. 1962년부터 시행되어 온 교육통계 사업은 400여 종의 국가 공식통계 중 사회분야의 핵심적인 통계사업으로 자리매김하고 있다.

교육통계연구센터를 포함하여 교육통계 정보를 제공하는 사이트는 [표 3-1-16]과 같다.

표 3-1-16 교육통계 제공 사이트

사이트명	통계정보	주소
교육통계 서비스	• 교육기본통계(유·초·중등, 고등, 평생교육통계) • OECD 교육통계 • 고등교육기관 졸업자 취업통계	kess,kedi,re.kr
대학알리미	• 대학 정보 공시 • 학생·교원 현황, 등록금 현황, 취업률 현황 • 학교법인 기본재산 현황	www.academyinfo.go.kr
학교알리미	• 초중고 학교정보 공시 • 학생·교원 현황, 재정·급식 현황, 학업성취	www.schoolinfo.go.kr
e-나라지표	• 학교급식실시 현황 • e-러닝 활용 현황 • 학자금 대출 현황 • 대학부설 평생교육원 현황 • 학점은행제 및 독학사 학위수여자 운영 현황	www.index.go.kr
학술정보통계 시스템	• 대학도서관 장서보유 현황, 예산 현황 • 대학 이러닝 보유 현황 • 대학·연구기관 보유 학술자원 현황	www.rinfo.kr

※ 출처: <http://www.moe.go.kr>, 교육부

나. 평생학습

평생학습은 형식적·비형식적 교육을 통해 직업 관련 역량 강화나 개인적 성장을 지속하는 학습 활동으로, 연령에 한정을 두지 않고 전 생애에 걸쳐 교육이 이뤄져야 한다는 교육관에 기초를 두고 있다.

교육과학기술부와 노동부가 공동으로 평생학습계좌제협의회를 구성하여 평생학습 계좌제(ALL, Accounts for Lilelong Learning)를 개발, 2008년부터 교육부가 중심이 되어 관련 교육훈련을 제공하고 있으며, 2016년 현재 국가평생교육진흥원이 교육부 위탁을 받아 이를 주관하고 있다.

평생학습에 대한 관리는 학습계좌를 통해 이뤄진다. 평생교육법 제23조에 따라 개인의 다양한 학습경험을 온라인 학습이력 관리시스템인 학습계좌에 기록·누적하여 체계적인 학습설계를 지원하고 학습결과를 학력이나 자격인정과 연계하거나 고용정보로 활용할 수 있도록 지원하고 있다. 평생학습 참여율 등 관련 정보는 교육통계 서비스 내 ‘주제별 교육통계 > 평생학습실태조사’를 통해 확인 가능하다.

그림 3-1-23 교육통계 서비스 내 평생학습정보 예시



※ 출처: <http://kess.kedi.re.kr/index>, 한국교육개발원

다. 직업능력개발

재직자와 구직활동을 하는 실업자를 대상으로 한 정부의 교육훈련 지원은 직업능력 개발사업을 통해 진행된다. 이는 근로자의 직무능력 향상을 위해 직업훈련 및 교육훈련을 실시하는 사업주와 자기능력개발을 위해 노력하는 근로자에게 훈련비용을 지원해 주는 사업으로서 사업주, 중소기업, 구직자(실업자), 재직근로자를 대상으로 직업능력개발 계좌제, 근로자 능력개발 카드제, 직업훈련 생계비 대부, 사업주 훈련비지원, 실업자 훈련비지원 등 다양한 형태로 이뤄지고 있다.

구직자 대상 직업능력개발 계좌제는 구직자의 자율적인 직업능력개발을 지원하는 제도이다. 구직자가 구직 등록과 계좌 신청 후 고용센터에서 상담을 진행하여 직업능력개발 훈련계좌 카드(내일배움카드)를 발급받을 수 있는데, 자신에게 적합한 훈련과정을 선택하여 수강하면 훈련비의 일부를 지원받을 수 있다. 실업자 훈련과정의 경우 주로 집체훈련과 인터넷 원격훈련으로 제공되고 있으며, 재직자 과정과는 다르게 집체훈련 중심으로 교육훈련 서비스가 제공되고 있다.



표 3-1-17 내일배움카드제 인터넷원격훈련 참가인원 현황

2013년	2014년	2015년
3,813명	3,305명	3,188명

※ 출처: 사업주단체 운영 결과 취합자료

재직자 대상 교육훈련 지원은 사업주 직업능력개발훈련으로 제공된다. 이는 고용보험에 가입한 사업주가 사전에 고용노동부장관으로부터 허가를 받아 교육훈련을 실시할 경우 교육훈련비를 지원받는 제도로, 자체 훈련시설에서 직접 실시하거나 타 훈련기관에 위탁하여 실시할 수 있으며, 집체훈련, 현장훈련, 원격훈련, 혼합훈련의 형태로 실시된다.

표 3-1-18 훈련방법별 훈련건수, 2008~2014년

(단위: 건)							
훈련 방법	2008년	2009년	2010년	2011년	2012년	2013년	2014년
집체훈련	1,552,126	1,460,254	1,531,493	1,521,774	1,642,815	2,213,744	1,960,005
	37.1%	34.5%	41.2%	50.2%	56.0%	66.4%	65.0%
현장훈련	2,834	4,017	2,809	2,260	2,301	2,712	4,075
	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%
우편원격	819,814	1,079,003	958,915	600,902	399,567	251,496	233,980
	19.6%	25.5%	25.8%	19.8%	13.6%	7.5%	7.8%
인터넷원격	1,795,074	1,681,784	1,218,838	896,409	872,885	842,677	798,741
	43.0%	39.7%	32.8%	29.6%	29.8%	25.3%	26.5%
혼합훈련	8,679	7,482	7,252	10,436	15,520	23,350	20,853
	0.2%	0.2%	0.2%	0.3%	0.5%	0.7%	0.7%
총계	4,178,527	4,232,540	3,719,307	3,031,781	2,933,088	3,333,979	3,017,654

※ 출처: HRD-NET

직업능력개발 훈련 프로그램 정보와 훈련과정별 취업률, 참여 및 수수료율, 중도탈락 현황 등 관련 정보는 직업훈련포털인 HRD-net(www.hrd.go.kr)을 통해 제공되고 있다.

라. 공공 학술정보 서비스

세계 학술정보 서비스 시장은 30조원으로 추산되며 그 가운데 국내 학술정보 서비스 시장은 2,000억원 규모로 추정된다. 학술정보는 학회, 대학, 민간연구소 등에서 발간한 저서, 논문, 연구 보고서를 수집/가공하여 DB로 구축한 후 교수, 학생, 연구자 등 필요한 사람들에게 무료 또는 유료로 제공되고 있다.

1) 미국 펌메드 센트럴

해외 공공 학술정보 서비스의 대표적인 예로 미국의 펌메드 센트럴(PMC, PubMed Central)을 들 수 있다. PMC는 미국국립보건원(NIH, National Institutes of Health)의 연구비 지원으로 산출된 연구 성과물을 누구나 이용할 수 있도록 구축한 생명공학 관련 전자 도서관으로, 미국국립보건원 산하 미국 국립의학도서관(NLM)에서 운영하고 있다. 연구비 지원을 받은 연구 성과 중 학술지 편집위원의 논문 심사를 통과한 최종본과 PMC 학술지표준에 부합되는 것만 수집하여 제 공함으로써 연구성과물 관리에 힘쓰고 있으며, 공공접근정책에 의거해 NIH 연구성과물을 전세계 누구나 이용할 수 있도록 하고 있다. PMC는 과학자들의 연구 성과에 대한 공공접근을 적시에 제공 함으로써 연구성과에 대한 가시성을 향상시켰다는 평가를 받고 있다.

그림 3-1-24 PMC 홈페이지

The screenshot shows the PMC homepage with the following structure:

- Header:** NCBI Resources, How To, Sign in to NCBI, Search bar.
- Main Banner:** PMC logo and description: "PubMed Central® (PMC) is a free full-text archive of biomedical and life sciences journal literature at the U.S. National Institutes of Health's National Library of Medicine (NIH/NLM)."
 - Get Started:**
 - PMC Overview
 - Users' Guide
 - Journal List
 - PMC FAQs
 - PMC Copyright Notice
 - Participate:**
 - Information for Publishers
 - Add a Journal to PMC
 - Participation Agreements and Options
 - FAQs for Publishers
 - Keep Up to Date:**
 - New in PMC | RSS
 - PMC Announce Mail List
 - Utilities Announce Mail List
 - Tagging Guidelines Mail List
- Central Statistics:**

3.9 MILLION Articles are archived in PMC.

Content provided in part by:

1892	307	3795
Full Participation Journals	NIH Portfolio Journals	Selective Deposit Journals
- Other Resources:**
 - PMC International
 - Open Access Subset
 - Eutilities
 - NLM LitArch
 - PMC Citation Search
- Public Access:**
 - Public Access and PMC
 - How Papers Get Into PMC
 - NIH Manuscript Submission System
 - My Bibliography
 - PMCID/PMID/NIHMSID Converter
- Footer:**

You are here: NCBI > Literature > PubMed Central (PMC) Write to the Help Desk

GETTING STARTED NCBI Education NCBI Help Manual NCBI Handbook Training & Tutorials Submit Data	RESOURCES Chemicals & Bioassays Data & Software DNA & RNA Domains & Structures Genes & Expression	POPULAR PubMed Bookshelf PubMed Central PubMed Health BLAST	FEATURED Genetic Testing Registry PubMed Health GenBank Reference Sequences Gene Expression Omnibus	NCBI INFORMATION About NCBI Research at NCBI NCBI News NCBI FTP Site NCBI on Facebook
--	---	---	---	---

※ 출처: <http://www.ncbi.nlm.nih.gov/pmc/>

2) 한국교육학술정보원 학술연구정보서비스(RISS)

한국교육학술정보원(KERIS)은 교육과 학술연구 분야에서 정보화를 추진하고 지원하는 교육부 산하의 공공기관으로 유치원부터 대학에 이르기까지 정보화 관련 사업을 추진하고 있다. 대표적인 서비스로 교육행정정보시스템 ‘나이스(www.neis.go.kr)’, 초·중·등 교육 정보 서비스인 ‘에듀넷(www.edunet.net)’, ‘학술연구정보서비스(www.riss.kr)’, 고등교육 교수학습자료 공동활용 서비스인 ‘KOCW(www.kocw.net)’, 해외박사학위논문 원문서비스인 ‘DDOD 서비스(ddod.riss.kr)’ 등이 있다.

대표적인 공공 학술정보 제공사이트인 RISS는 국내·외 대학 및 연구기관이 생산하는 다양한 학술 콘텐츠를 수집·유통해 연구자에게 필요한 서지정보와 원문을 검색하고, 관련 정보를 다운로드 받아 활용할 수 있도록 지원하는 사이트로, 1998년 5월부터 서비스가 시작됐다. 2016년 현재 학위논문, 국내학술지논문, 해외학술지논문, 학술지, 단행본, 공개강의 서비스를 제공하고 있다. 그 외 공공영역에서는 한국연구재단, 한국과학기술정보연구원, 한국과학기술단체총연합회 등에서 관련 서비스를 제공하고 있다.

3. 교육훈련정보 시장 동향

가. 개요

2008년 산업자원부는 지식기반서비스산업 육성정책을 수립하고 유망 지식서비스 업종으로 이러닝을 선정한 바 있다. 이는 제조업의 발전을 지원하고 높은 고용흡수력을 통해 고용창출에 기여할 것으로 판단되었기 때문이다. 이러닝(전자학습)산업발전법에 의하면 이러닝은 전자적 수단, 정보통신 및 전파방송 기술을 활용하여 이루어지는 학습이다. 즉 전자적 방법을 이용하여 학습 서비스를 제공하는 사용자 중심의 학습 패러다임을 의미하며, 학습자의 자기주도적 학습을 가능하게 하는 새로운 학습 유형으로 기존의 면대면 교육의 대안 또는 보조 수단으로 각광받고 있다. 이러닝은 집체훈련과의 결합을 통해 블렌디드 러닝, 플립드러닝(flipped learning) 등으로 발전하고 있다.

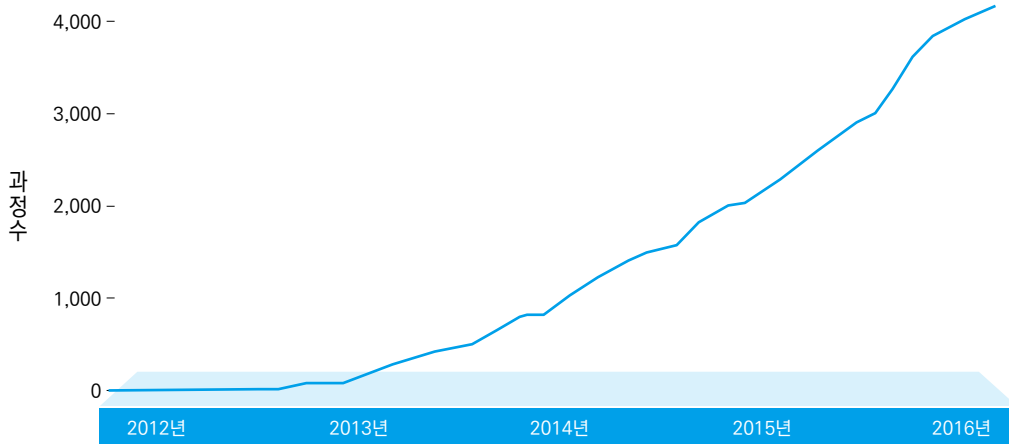
나. 온라인 공개수업 확산

1) 온라인 공개수업(MOOC)

최근 온라인 공개수업(MOOC)이 빠른 속도로 대중적인 인기를 끌고 있다. MOOC(Massive Open Online Course)은 대학 수업을 온라인으로 접속해 누구나 들을 수 있도록 하는 서비스이다. MOOC 이전에도 OCW처럼 무료로 대학 강의를 온라인으로 제공하는 서비스가 있었다. 그러나 학습자가 수동적으로 듣기만 하던 온라인 강의와는 달리 MOOC에서는 교수자-학습자, 학습자

-학습자간 다양한 상호작용이 가능하며, 수강인원의 제한 없이 누구나 강의를 들을 수 있고 다양한 지식과 경험을 가진 전세계 학습자가 지식을 공유할 수 있어 새로운 학습경험을 제공해 준다.

그림 3-1-25 MOOC에 공개된 강의 수



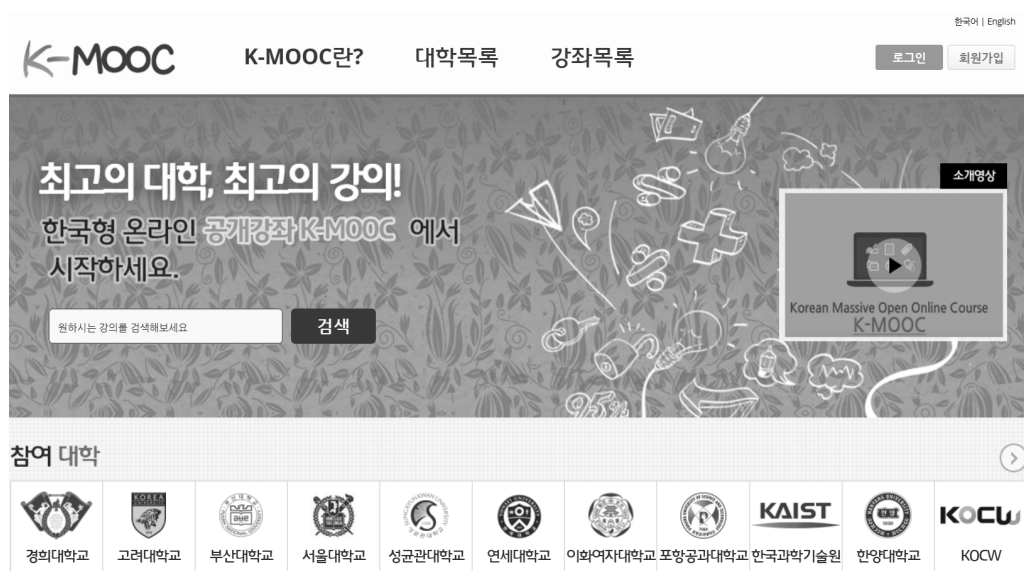
※ 출처: 블로터넷이 인용한 클래스센트럴블로그(<https://www.bloter.net/archives/246853>)

1세대 MOOC는 하버드, MIT, 스탠포드대학 등에서 진행된 강의를 녹화해 온라인으로 제공하는 방식으로 서비스가 이뤄졌으며, 대표적으로 유다시티(www.udacity.com), 코세라(www.coursera.org), 에드엑스(www.edx.org)를 들 수 있다. 유다시티는 2011년 스탠포드대학의 컴퓨터공학 수업을 무료로 제공한 것에서 출발한 영리 교육기관으로 스탠포드대 출신 교수들이 직접 강의하는 형태로 서비스가 제공되고 있다. 초기에는 무료 강의를 제공했으나 현재는 유료 강의 비중을 높혀가고 있다. 코세라는 2012년 스탠포드대학의 컴퓨터공학과 교수인 앤드류 응(Andrew Ng)과 대프니 콜레(Daphne Koller)가 공동 창업한 영리기업으로 컴퓨터공학뿐만 아니라 과학, 예술, 인문, 비즈니스, 수학 등 다양한 강의를 제공하고 있으며 예일대, 카이스트 등 전세계 우수 대학이 참여하고 있다. 콘텐츠 수, 방문자 수에서 가장 큰 규모를 자랑하는 코세라는 1,300개가 넘는 강좌를 개설하고 있다. 에드엑스는 MIT와 하버드대가 공동투자한 기관으로 유다시티나 코세라와 달리 비영리를 추구하고 있다. 최근 2세대 MOOC 등 다양한 MOOC 플랫폼이 등장하고 있으며, 유료 강의 비중이 증가하고 있다.

2) 한국형 무크

한국형 무크(K-MOOC)는 MOOC에 적극 동참하고 국내 대학의 최고 수준의 강의 공개를 통해 좋은 강의를 확산함으로써 대학 수업의 질적 제고를 위해 도입되었다. K-MOOC 사업의 목적은 글로벌 명품 콘텐츠 개발 및 공개를 통한 고등교육의 실질적 균등 실현, 대학의 자율성과 다양성이 존중되는 공용 플랫폼 구축 운영, 한국형 MOOC 활성화를 위한 기반 조성으로 요약된다. 2015년 K-MOOC 참여대학으로 서울대 등 총 10개교를 선정하여 콘텐츠를 개발 후 서비스를 제공하고 있다. 과정은 대학에 개설되었거나 개설 예정인 정규교과로, 강의는 한국어로 진행되며 영어 자막을 함께 제공하고 있다. 수강생은 각 대학이 정하는 기준을 충족한 경우 대학 명의의 이수증(Certificate)을 발급받을 수 있다. 이처럼 대학 간 경계를 넘어 우수한 강의를 온라인을 통해 폭넓게 제공함으로써 양질의 교육에 대한 기회 균등을 실현할 것으로 기대된다.

그림 3-1-26 K-MOOC 홈페이지



※ 출처: <http://www.kmooc.kr>

4. 향후 전망

전세계적인 불황과 고령화 문제, 취업난 및 비정규직 증가와 같은 고용구조 변화 속에서 세계 각국은 국가인적자원의 경쟁력 강화를 위해 많은 노력을 기울이고 있다. 공공정보에 대한 접근성을 높이기 위해 공공영역에서는 학술정보와 각종 통계정보를 투명하게 공개하고 있으며, 초중고



및 대학 등 정규교육기관과 평생학습기관에서도 양질의 교육 콘텐츠를 무료로 제공하고 있다. 일 예로 OCW나 MOOC를 통해 대학이 보유한 교육 콘텐츠를 제공함으로써 특정 대학이라는 물리적 울타리에 구애되지 않고 온라인을 통해 누구나 양질의 교육을 받을 수 있게 되면서 학교교육 및 평생 교육의 경쟁력이 향상될 것으로 기대된다.

ICT 발전에 따라 집체 중심의 교육훈련은 이러닝이나 스마트러닝 등 원격훈련으로 대거 전환되어 왔으며, 집체훈련과 인터넷 원격훈련의 장점을 모은 블랜디드 러닝과 플립드러닝 등 진화한 교육훈련 방식이 교육현장에서 뿌리내리고 있다. 앞으로 네트워크와 스마트 디바이스의 기술적 진화, 빅데이터 처리기술 및 기계학습의 발전에 따라 개인화된 콘텐츠가 구현될 것으로 기대하며, 증강현실(AR)이나 가상현실(VR) 기술을 기반으로 높은 몰입도의 체험기반 학습이 확산될 것으로 예상된다.



제2장 데이터 솔루션 동향

제1절 데이터 수집

1. 개요

“우리는 절대로 데이터를 내다버리지 않는다.” 데이터를 갖고 있는 기업이 결국 다양한 사업 기회를 갖고 성공할 수 있다는 확신에 찬 아마존 창업주 제프 베조스의 말이다.

우리는 이미 빅데이터 시대에 살고 있다. 빅데이터란 기존 데이터베이스 관리도구로 데이터를 수집·저장·관리·분석할 수 있는 역량을 넘어서는 대량의 정형 또는 비정형 데이터 집합을 의미하는데, 이 데이터를 활용하기 위한 다양한 분석 방법들과 서비스를 구성하기 위한 기술들이 넘쳐나고 있다.

그러나 아무리 좋은 분석 방법과 기술이 있다고 해도 양질의 데이터를 확보하지 못한다면 제대로 된 분석 품질과 서비스를 기대하기 어렵다. 국내기업들이 빅데이터를 다루는 역량이 부족한 것에 비해 구글이나 아마존 같은 기업은 빅데이터 시대의 강자들로, 이들은 모두 데이터를 중시하는 문화를 갖고 있다.

구글이 강력한 웹검색 서비스를 할 수 있는 것은 좋은 랭킹 알고리즘, 빠른 검색 속도와 같은 검색 관련 기술이 좋기 때문이기도 하지만 가장 많은 웹 페이지를 수집하여 검색 대상을 많이 확보하고 있다는 것이 제일 중요한 부분이다.

빅데이터에는 일반 웹사이트 온라인 문서, 소셜 미디어, 기업내부자료, CRM 등 정형/비정형 데이터를 포함한 여러 종류가 있으며, 이를 수집하기 위한 기술도 다양하다. 본 절에서는 웹사이트를 수집하기 위한 웹 크롤러 솔루션의 기술적 동향을 살펴해보도록 한다.

2. 제품 동향

웹 크롤러는 조직적, 자동화된 방법으로 월드와이드웹을 탐색하는 프로그램으로, 앗츠(ants), 자동 인덱서(automatic indexers), 봇(bots), 웜(worms), 웹 스파이더(web spider), 웹 로봇(web robot) 등의 이름으로 불리고 있다. 웹 크롤러는 시드(Seed)라고 불리는 첫 번째 웹 페이지에 방문하여 해당 페이지를 스크랩하고 페이지의 모든 하이퍼링크를 추출하여 저장한 뒤, 새로운 URL 하나를 꺼내 앞의 과정을 반복하는 비교적 간단한 구조를 갖는다. 이런 간단한 구조로 인하여 수집 솔루션을 만드는 것은 쉬운 편으로 다양한 수집 솔루션이 국내외에 출시되어 있다. 또한 일정 수준 이상의 오픈소스 기반 수집엔진도 많이 등장하고 있다. 안정적인 운영과 지속적인 개선 등을 원하는 기업 및 공공기관 등에서 공인시험 기관으로부터 인증받은 제품을 주로 도입하고 있다.

하지만 인증받은 제품이라 할지라도 특수 목적에 맞게 개발되어 수집이 제한적이어서 범용 제품으로 불리기에 부적합한 제품들도 있다. 범용 제품이라면 일반 사용자가 수집 설정하기 용이하고 다양한 형태의 웹 페이지를 수집할 수 있어야하며, 사용자가 원하는 형태의 데이터 포맷(혹은 DB)으로 변환이 용이해야 한다.

최근에는 HTTP 프로토콜을 통해 오픈 API를 제공하는 사이트의 수집 요구가 증가하고 있으며, 범용 제품에서 오픈 API 설정까지 할 수 있는 제품 등이 개발되고 있는 추세이다.

3. 기술 동향

폭발적으로 증가하고 있는 웹 페이지는 다양한 형태가 존재하며, HTML 제작의 용이성과 자바 스크립트의 난개발로 인하여 웹 크롤러는 다양한 기술적 이슈들을 갖고 있다.

웹 페이지를 수집하는 데 있어 동일 페이지를 재방문하거나 반복·수집하는 것을 막기 위한 방법은 필수 기능이다. 일반적으로 사용하는 방법은 이미 방문했던 URL은 재방문하지 않는 것이다. 하지만 웹 페이지의 URL이 달라지는 경우가 있는데, 파라미터가 세션 id나 현재 시간, 페이지 번호 등으로 구성되어 매번 다른 웹 페이지로 인식될 수밖에 없다. 따라서 기본적인 중복 체크는 URL로 하되, 제목이나 작성날짜, 본문 등으로 중복 체크를 설정할 수 있어야 한다.

웹 페이지에는 수집해서 활용하고 싶은 콘텐츠만 포함되어 있지 않고, 광고글이나 원하지 않는 하이퍼링크 등이 혼재되어 있다. 예를 들어 신문사 사이트에 들어가면 하나의 웹 페이지는 해당 기사의 텍스트 외에도 필요 없는 광고글로 가득한 것을 볼 수 있다. 필요한 영역의 데이터만 잘 가져오기 위해 다양한 연구가 진행되고 있지만, 범용 제품은 한 사이트만을 대상으로 수집하는 것이 아니기 때문에 개발에 어려움이 따른다. 초창기 크롤러에서부터 이 문제를 풀기 위해 Wrapper induction을 포함한 지식기반 학습 방법론 등의 연구들이 있으나 타깃으로 삼은 특정 웹사이트

몇 개 정도만 적용이 가능한 수준이다.

HTML5가 등장하기 전까지 자바스크립트는 동적인 페이지를 구성하고 다양한 표현을 하기에 손쉬운 방법을 제공하였다. 하지만 허술한 구조와 HTML이 갖는 범용성이 더해져 브라우저로 웹 사이트에 방문할 때 많은 오류들을 접하게 되었다. 이 문제는 웹 크롤러 입장에서도 자바스크립트가 거대한 벽이 아닐 수 없는데, URL이 자바스크립트로 작성된 경우에는 해당 평선을 호출해서 고유의 URL을 얻어올 수 있어야 한다. 이 문제에 많은 크롤러들은 사람의 손을 빌려 사이트별로 평선이 호출된 뒤 얻어올 URL을 미리 확보하여 파라미터만 바꾸는 식으로 처리하곤 하는데, 다양한 사이트를 수집해야 한다면 엄청난 노력이 요구된다. 솔루션으로써의 웹 크롤러라면 가상 브라우저 등을 탑재하여 대부분의 자바스크립트로부터 URL을 자동으로 추출할 수 있어야 한다.

기존의 HTML4 기반의 웹 페이지들은 페이지네이션(Pagination), 페이지 스트럭처(Page Structure), 그리고 인터널 서치 페이지(Internal Search Page) 등과 같은 문제들을 안고 있었으며 특히, 에이잭스(AJAX)나 앵귤러JS(AngularJS)와 같은 자바스크립트(JavaScript) 기반의 URL 변경 없이 페이지 콘텐츠를 부분적으로 변경하는 기술들이 널리 이용되는 바람에 웹 크롤러 수집 능력에 한계가 있을 수밖에 없다.

이에 비해 HTML5에서는 푸시스테이트(pushState) 메소드를 사용하여 에이잭스나 앵귤러JS를 사용하지 않고도 제이쿼리(Jquery)만을 이용하여 페이지 내용을 부분적으로 바꿀 수 있게 되었다. 이뿐만 아니라 URL까지 변경이 가능하게 되어 HTML5를 고려한 웹 페이지는 기존의 크롤러 기술로도 쉽게 동적으로 변하는 페이지를 수집할 수 있게 되었다. 또한 위에서 언급된 페이지네이션, 페이지 스트럭처 및 인터널 서치 페이지 문제 또한 간단하게 해결이 가능하게 되었다.

그러나 아직 대부분의 웹페이지가 HTML4까지만 지원하고 있기 때문에 크롤링에 많은 어려움이 따른다. 이를 해결하기 위해서는 모든 페이지를 방문할 때마다 자바스크립트를 실행하면서 수집해야 하는데, 이는 사람이 직접 방문하는 것과 크게 다르지 않을 정도로 속도가 느려지는 문제가 있다. 따라서 대부분의 크롤러들은 에이잭스를 사용하는 HTML4 기반의 페이지를 제대로 수집하지 못하고 있는 상태이다. 그리고 현실은 대부분의 웹 페이지는 아직 HTML4 기반에 머무르고 있다.

4. 시장 동향

웹 크롤러는 다양한 수집 목적으로 개발되었으며 대부분은 특정 서비스에 맞춰 사용되고 있다. 솔루션으로서의 웹 크롤러는 기본적인 기능 외에 범용성과 안정성이 높아야 한다.

주요 기능을 살펴보면 수집물 관리가 편리해야 하며, 링크가 유효한지 체크할 수 있는 데드링크 관리 기능, 중복 문서 처리와 패턴기반 데이터 필터링을 통한 클렌징 기능, 다양한 DB와 연동되며

첨부파일, 이미지 등을 수집할 수 있는 확장성, 그밖에 유연한 자바스크립트 처리, 인증페이지 수집, 다국어 지원, SNS 수집 등이 있다.

최근에는 오픈 API를 통해 정보를 제공하고 있는 사이트들이 많이 늘어나고 있다. 통계청, 기상청 같은 공공 기관에서 HTTP 프로토콜을 이용한 오픈 API를 외부에 제공하고 있으며, 이들 반정형 데이터에 대한 분석 요구가 증가하고 있어 일반 사용자 수준에서 이러한 데이터를 손쉽게 수집하기 위한 수집기 개발이 필요해지고 있다. 주요 기능으로는 파라미터를 설정할 수 있는 화면과 반복되는 호출 방식 지원 및 수집 결과를 파싱하기 위한 설정 등이 있다.

이와 같이 현재 시장에서 요구되는 필수 기능은 다양한 수집 기술을 필요로 하며 점차 증가하고 있는 추세이다. 이는 새로운 기술의 웹 페이지가 꾸준히 늘어나고 있으며 새로운 서비스 출시로 인한 수집 솔루션 요구가 꾸준하다는 것을 방증한다.

수집/검색 시장 규모를 통해 간접적으로 수집 시장의 규모를 확인해 볼 수 있다. 2015년 국내 수집/검색 부분 시장 규모는 약 1,293억원 정도이며, 전체 시장의 약 77% 정도를 국내 기업이 점유하고 있어 타 분야에 비해 국산 제품이 강세를 나타내고 있다.

5. 향후 전망

최근 외부 비정형데이터와 기업 내부의 데이터를 융합하여 분석하려는 기업이 많이 늘어나고 있는 추세이다. 외부의 비정형데이터는 폭발적으로 그 수가 증가하고 있는데, 분석 기술의 발전으로 데이터가 많을수록 의미 있는 분석결과를 보여주어 외부 데이터를 잘 수집하고자 하는 요구는 갈수록 많아지고 있다.

현재 전세계의 웹사이트 수는 10억개를 넘어섰는데, 그 안에 있는 웹 페이지 수는 수조개에 이른다. 폭발적으로 증가하는 웹 페이지에서 적절한 양질의 데이터만 크롤링하는 것은 좋은 분석결과를 얻기 위해 꼭 필요한 일이다.

수집 솔루션들은 빠른 수집 속도와 다양한 데이터 형태를 수집하기 위한 옵션들이 많이 갖추어져 있다. 또한 국내 웹 페이지들은 난무하는 자바스크립트와 액티브X의 사용 등으로 해외 웹 페이지들에 비해 수집하기가 까다로운데, 이러한 페이지들을 수집하기 위해 발전해 온 국산 수집 솔루션들은 그 우수성이 입증되어 있다고 볼 수 있다.

오픈 API를 통해 공공데이터를 제공하는 사이트가 갈수록 많아지는 가운데, 수집 솔루션에서 오픈 API를 수집하고자 하는 요구가 있어 기존 솔루션에서 이 부분을 어떻게 가져갈 것인가가 중요한 차별화 요소가 될 것으로 전망된다.

제2절 데이터 설계

1. 개요

기업에서 요구하는 시스템 구축을 위해 정보 시스템 태동 이후 많은 학자들은 프로세스 중심의 정보시스템 분석, 설계 기법만으로는 시스템의 문제점들을 해결할 수 없다고 느끼고, 정보시스템의 핵심은 기업이 가지고 있는 자산(데이터)을 어떻게 정확하게 관리하여 결과를 오류 없이 보여줄 수 있는지에 대한 부분을 해결해야만 한다는 것을 인지하게 되었다.

이런 문제의식은 프로세스 설계나 개발보다는 정확한 업무과약을 통해 현실 속에서 정보간의 관계를 이해해야 한다는 결론에 이르고, 실세계를 잘 표현할 수 있는 관계형 데이터베이스나 개체 관계 모델링 기법을 발전시켜 데이터 설계 분야의 부흥을 이끌어 왔다.

데이터 설계는 데이터 모델의 이해를 필요로 하는 그룹들(최종 사용자, 시스템 분석가, 데이터 베이스 관리자, 타 프로젝트에서 작업하는 분석가)에게 요구사항이 명확하게 공표됨은 물론 최종 사용자가 분명하게 파악할 수 있는 수준으로 작성되어야 한다. 정확한 데이터 설계는 시스템을 구축하는 애플리케이션 개발자에게도 중요한 기능(올바른 개발 방향)을 제공하여 고품질의 산출물이 나올 수 있도록 해야 한다.

데이터 설계는 프로세서, 데이터의 설계라는 기본 위에 명칭·정의·형식·규칙 등의 표준을 정하고, 표준에 맞는 설계를 진행함으로써 시스템의 품질을 높리게 되는데, 이를 지원하는 다양한 제품들이 있다.

2. 제품 동향

현재 국내 데이터베이스 모델링 제품은 외산 제품이 국내 제품에 비해 높은 점유율을 차지하고 있는 것으로 나타난다.

주로 사용되는 제품으로는 CA테크놀로지스 ERwin, 엠바카데로 ER/Studio, SQL Power Group SQL Power Architect 등의 외산 제품이 있으며, 국산 제품으로는 엔코아 DA#, 토마토시스템 eXERD가 있다.

ERwin은 r9.1 이후 직관적 그래픽 인터페이스를 통해 엔터프라이즈 데이터를 관리할 수 있는 데이터 모델링 협업 환경을 지원한다. 모델링의 편의를 위한 다양한 고급 기능이 추가됐으며, 클라우드 DB인 SQL 애저를 지원하고 DB 클라이언트나 드라이버가 없어도 레파지토리에 빠르게 접근할 수 있는 웹 기반 3터어 방식을 지원한다. 또한 서드파티 형식으로만 지원되던 웹 뷰어를 클라이언트 뷰어인 네비게이터와 웹 포털로 직접 지원한다.

2014년 3월 엠바카데로는 CA테크놀로지스의 ERwin을 인수하려 했으나 데이터모델링 툴 시장의 독점을 우려하여 미국 법무부에서 이를 기각하였다. ERwin의 인수가 무산되자 엠바카데로는 ER/Studio의 기능 개선에 대한 지원을 강화하고 있다. ER/Studio는 엔터프라이즈 데이터모델링 도구로 비즈니스 모델링 및 UML(Unified Modeling Language) 모델링을 포함하는 엔터프라이즈 아키텍처 모델링을 지원하며, 모델 및 오브젝트를 확인·관리할 수 있는 웹 포탈도 지원하고 있다. V9.7버전에서는 SQL 서버 2014, 몽고DB(MongoDB), 하둡 하이브(Hadoop Hive) 등의 DB 지원을 추가했고, v10.0 버전에서는 64비트 지원 및 레파지토리 처리속도를 향상시키는 등 지속적인 기능 개선이 이루어지고 있다.

DA#은 개괄 모델부터 개념, 논리, 물리 모델까지 데이터 아키텍처 구축의 전 단계를 총괄적으로 지원하는 전사 데이터 아키텍처 모델링 도구이다. DB모델링 제품 중 유일하게 거의 모든 국산 DBMS를 지원하고 있다. 또한 거버넌스 솔루션인 데이터웨어 제품들과의 연계를 통해 전사 거버넌스 체계를 수립하고 관리할 수 있도록 쉽게 기능을 확장할 수 있는 것이 특징이다.

eXERD는 이클립스 기반의 모델링 도구로 전문가가 아닌 일반 개발자들도 쉽게 데이터 모델링을 할 수 있도록 사용자 편의성을 향상 시킨 국산 모델링 제품으로 논리/물리 모델을 동시에 보면서 설계할 수 있으며, 다양한 자동화 기능과 인터랙티브한 애니메이션 효과 등이 특징인 제품이다.

3. 기술 동향

초창기 모델링 도구는 최종적으로 관계형 DB에 전이할 목적으로 정보 위상을 설계하기 위한 도구였으며, 각 단계별 위상을 편집하고 다음 단계로 전이하는 것을 돕는 것이 주목적이었다. 이 과정을 용이하게 하기 위해 용어사전, 도메인 등과 같은 정보 모델에 적용 가능한 정책들을 쉽게 운용할 수 있는 형태로 발전하였다.

빅데이터에 대한 중요성이 부각되며 데이터 규모가 방대해짐에 따라, 정보 모델의 변경 이력 관리, 권한을 통한 협업 지원 등의 필요성이 부각되었고, 이러한 환경을 위해 추가적인 도구가 공급되어 모델링 도구와 연동 되거나, 모델링 도구 자체에 이러한 기능을 포함하는 형태로 발전하고 있는 추세이다.

또한 클라우드 DB 지원 등 변화하는 IT환경에 맞춰 단순히 데이터 아키텍처 모델링 도구가 아닌 전사 아키텍처 모델링 솔루션으로 발전해 나가고 있다.

4. 시장 동향

국내 데이터베이스 모델링 시장은 빅데이터 이슈와 맞물려 최근 5년간 연평균 9.1%의 성장을 보이고 있다.

특히 2014년 약 200억원 규모를 돌파하며, 전년 대비 21.7%의 성장률을 보이며 큰 폭으로 성장하고 있다. 데이터 모델링 시장의 높은 성장은 빅데이터 이슈와 맞물려 공공·금융·기업들의 아키텍처 수요가 급증했기 때문이라고 풀이된다.

그 동안 국내 빅데이터 시장은 소셜 분석 등 대용량의 비정형 데이터들에 대한 분석에 초점을 맞췄으나, 최근 대기업 및 선두 기업들을 중심으로 빅데이터 분석뿐 아니라 플랫폼, 서비스적으로 접근하기 위한 테스트가 진행되고 있다. 이에 따라 새로운 플랫폼의 확산, 빅데이터를 대상으로 하는 DB모델링에 대한 수요 증가로 인해 DB모델링 시장은 더욱 확대될 전망이다.

한편 국내 모델링 시장은 ERwin을 비롯한 외산 제품들이 국내 데이터 모델링 시장을 장악하고 있다. 하지만 국내 기업들의 적극적인 마케팅과 함께, 정부 및 공공기관의 국산 소프트웨어에 대한 인식이 높아지면서 관련 육성 정책이 지속적으로 추진됨에 따라 국산 모델링 제품의 도입이 늘어나고 있다.

5. 향후 전망

데이터베이스 모델링 도구는 데이터 설계·관리 컨설팅과 시스템 구축 방법론으로 분석, 설계하여 데이터를 적절히 표현하는 데 필요한 소프트웨어이다. 시스템의 안정성과 유연성, 성능 등에 미치는 영향이 크고, 고품질의 데이터 모델을 확보하기 위한 시스템 개발에 있어 가장 핵심적인 과정이라는 인식 때문에 그 수요는 지속될 것으로 예상된다.

빅데이터 분석 및 활용에 대한 관심이 고조되고 관련한 기업의 투자가 증가하면서 정보 분석과 관련한 사업모델이 늘어나며 데이터베이스 모델링 시장도 지속적으로 성장할 것으로 전망된다. 또한 기업의 IT시스템이 복잡해짐에 따라 전사적 정보관리 체계가 더욱 필요하게 되며, 데이터를 관리하기 위한 데이터 모델링 도구로 활용되는 사례가 늘어날 전망이다.

향후 새로운 IT 환경에 맞춰 클라우드 서비스와 빅데이터 접목도 활발해지면서 데이터베이스 모델링 도구들도 새로운 환경을 지원하기 위해 지속적으로 발전할 것으로 예상된다.

제3절 DBMS

1. 개요

데이터베이스관리시스템(DBMS, Database Management System)은 사용자 또는 응용프로그램이 데이터베이스에 데이터를 저장 및 수정하거나 데이터베이스 내의 데이터를 사용하는 데 안전한 접근을 보장해 주며 데이터베이스 내 저장된 데이터들을 위한 다양한 기능을 제공하는 소프트웨어의 집합이다.

하나의 응용 프로그램이 운영에 필요한 데이터를 저장하기 위해 자신이 관리하는 저장소(파일)를 가지고 있었으나 응용 프로그램마다 중복된 저장소를 가지며 발생하는 효율성과 일관된 정보관리의 어려움을 개선하기 위해 하나의 저장소를 여러 응용 프로그램이 활용하는 형태로 발전하였다. 저장소에 저장된 데이터를 일관되게 관리하기 위해 데이터베이스 개념이 생겨났으며, 이 데이터베이스를 관리하기 위한 시스템인 DBMS를 사용하게 됨으로써 다수의 응용 프로그램들은 하나의 데이터베이스를 서로 공유하며 사용할 수 있게 되었다. 이를 위해 DBMS는 데이터베이스의 구성부터 접근 방법, 데이터에 대한 안전함(무결성) 및 관리에 대한 모든 책임을 가진다.

하나의 응용 프로그램에 종속되어 저장되던 저장소(파일 시스템)의 사용으로 발생할 수밖에 없었던 데이터의 중복성 해결을 위해 원자성(Atomicity), 일관성(Consistency), 고립성(Isolation), 지속성(Durability)을 지원하며 등장한 DBMS는 계층형(Hierarchical) DBMS, 네트워크 DBMS 등으로 발전해왔다.

현재까지 가장 많이 사용되고 있는 DBMS는 관계형(Relational) DBMS로 IBM 산호세 연구소의 에드거 F. 커드가 도입한 관계형 모델을 기반으로 하여 만들어졌으며, 현재 사용되는 대부분 데이터베이스의 근간이 되었다. 최근에는 객체 개념을 반영한 객체관계형(Object-Relational) DBMS, 객체지향형(Object-Oriented) DBMS 등도 개발되었다.

1970년대 후반 등장한 관계형 DBMS가 40년이 지난 오늘날까지도 활용될 수 있었던 것은 SQL(Simple Query Language)의 일반화 때문이다. SQL이 응용 프로그램에서 데이터에 접근하기 위한 표준 언어로 자리잡으며, SQL을 통한 데이터 처리 방법이 사용자 및 응용 프로그램 개발자, 시스템 관리자들에게 이미 친숙해졌기 때문이다.

그러나 데이터를 생성하는 주체가 사람이 아닌 머신(machine) 또는 기기(Device) 등으로 확대되며 새로운 형식의 머신 또는 기기에서 생성되는 로그 데이터가 출현하였으며, 스마트 기기 및 CCTV 등의 확대로 사진, 영상 데이터 저장 및 활용을 위해 기존의 관계형 DBMS가 중심을 이루던 데이터 처리 기술에도 변화를 불러왔다.

특히 소셜 네트워크 서비스의 폭발적 성장에 따라 SNS 등에서 발생하는 데이터를 저장·활용하기 위한 NoSQL(Not Only SQL)도 주목받고 있다. 이는 기존의 DBMS를 사용하여 SNS 데이터를

처리하기에는 비용이 효율적이지 못하며, 기업에서 관리하던 데이터 유형이 아닌 새로운 유형의 데이터에 대한 솔루션으로 관계형 DBMS는 무거웠기 때문이다.

하지만 NoSQL 기술을 적용하기에 적합한 부분이 있기 때문에 대다수 기업에서는 NoSQL 기술과 기존 관계형 DBMS를 복합적으로 활용하고 있다. NoSQL은 SNS에서와 같이 정형화되지 않은 데이터들을 저장하고 활용하는데 사용하고 있으며, 기존 업무용 데이터와 결합하여 활용하고 싶은 경우 관계형 DBMS로 데이터를 적재해 통합 분석하는 시도들도 이뤄지고 있다.

최근에는 반복 생산되는 머신데이터나 스트림 형태로 생산되는 데이터들을 적재 후 분석하는 처리방식 대신 실시간으로 데이터를 처리하는 CEP(Complex Event Processing) 방식도 주목을 받고 있다.

2. 제품 동향

DBMS는 소프트웨어 중에서도 가장 기술적인 난이도가 높은 기반 소프트웨어로, 시스템에서 매우 중요한 위치를 차지하고 있어 실제 IT 시스템에 적용된 사례를 가진 DBMS는 많지 않다. 우리나라를 비롯해 미국, 독일 등 소수 국가들만이 실제 상용 DBMS 제품을 개발하였다.

이는 DBMS 기술이 다양한 애플리케이션과 연동되어 운영되어야 하며, 데이터를 안전하게 보관하는 것뿐 아니라, 시스템 운영 중 발생할 수 있는 여러 가지 장애상황에서도 데이터의 무결성을 보장해야 하는 요구를 받기 때문이다. 특히 미션 크리티컬한 업무에 적용되는 메인 DBMS는 무중단 운영, 대량의 사용자 접근에도 안전한 데이터 처리 및 대량 데이터를 처리하는 등 다양한 상황에서의 DBMS 기능을 요구하고 있다.

기업에서 활용할 수 있는 DBMS를 개발하는 것은 지속적인 투자와 수십년의 장기간 연구개발이 수반되어야 하므로, 진입 장벽이 매우 높은 시장이다. 기존 고객이 선호하는 기업들의 제품을 대체하고 선택받는 것은 더더욱 어렵다.

데이터베이스를 관리하는 DBMS는 상위로 다양한 데이터베이스 관련 서드파티 솔루션과 연계 및 응용 프로그램이 위치하며 생태계가 구성되어 있다. 새롭게 응용 프로그램을 개발하거나 시스템을 구성할 때 사용자는 기존에 사용하거나 익숙한 DBMS 제품을 선택할 수밖에 없는 경우가 많아 타 DBMS를 선택하는 경우 선택의 폭이 넓지 않아 종속적 구조가 지속되는 특성이 있다.

현재 국내에 DBMS 제품을 개발·공급하는 기업으로는 오라클, IBM, 마이크로소프트, SAP 등의 글로벌 기업과 티맥스데이터, 알티베이스, 큐브리드, 선재소프트, 리얼타임테크 등 국내 기업이 있다.

오라클은 2013년 7월 ‘오라클 데이터베이스 12c’를 출시한 후 클라우드에 초점을 맞추고 있으며, DB 어플라이언스 역시 매년 업그레이드해 시장을 공략하고 있다. 특히 엑스데이터 X6-2와

오라클 데이터베이스 어플라이언스(ODA) 판매에 집중하고 있다. 또한 오라클은 자사 클라우드에서도 클라우드 데이터베이스 서비스를 제공하고 있다.

IBM은 2016년 7월 IBM DB2 11.1를 출시하며 실시간 지원과 인메모리 기술 지원을 제공하고 있다. DB2 11.1은 기존 버전인 9.7에서 11.1로 곧바로 업그레이드를 제공하며, 다중 작업을 지원한다. 또한 모바일, 고급 분석, 코그너티브, 고가용성 트랜잭션을 지원한다. IBM 역시 클라우드를 지원함으로써 온프레미스에서 클라우드로 이전하여도 기업 운영에 영향을 주지 않게 하고 있다.

마이크로소프트는 2016년 4월 SQL Server 2016을 출시하였다. 기존 온프레미스는 물론 클라우드 상에서도 사용 가능한 데이터베이스로 가장 최근 발표된 제품이다. SQL Server 2016의 가장 큰 특징은 고객이 데이터 관리와 분석에 필요한 기능을 모두 제공하고 있다는 점이다. 이로써 OLTP(Online transaction processing) 및 데이터웨어하우스 등의 다양한 업무환경에서 SQL 서버를 효과적으로 활용할 수 있다.

SAP은 2015년 11월 SAP S/4HANA를 출시하며 차세대 비즈니스 스위트를 통한 엔터프라이즈 소프트웨어 플랫폼을 강화하고 있다. 이는 고객사의 디지털 변혁을 견인하기 위한 차세대 비즈니스 스위트를 기반으로 하고 있다. 레노버, IBM, HP, 삼성전자 등의 다양한 하드웨어 벤더와 동맹을 통해 'SAP HANA'를 중심으로 협력을 주도하고 있으며, 클라우드 플랫폼인 'SAP HANA Vora' 출시했다.

티맥스테이터는 2015년도 3월 대용량 DB를 목표로 한 Tiberio 6를 출시하였다. Tiberio 6는 타 DBMS와 호환이 가능하며, DB 전환시 영향도를 사전에 예측하고 자동 전환하는 T-up 기능을 제공한다. 최근 새로운 DB 어플라이언스인 제타데이터(ZetaData)를 글로벌 하드웨어 기반으로 출시하였고, AWS 및 KT 클라우드에서 Tiberio 서비스를 출시하며 클라우드 시장까지 확대하고 있다.

알티베이스는 2015년 6월 기존보다 성능을 높인 Altibase HDB 6.5를 출시하며 시장 공략을 강화하고 있다. Altibase HDB 6.5는 인텔 제온 E7 장비에서 기존 HDB 6.3보다 1.5배의 성능을 보일 정도로 성능 향상을 보이고 있으며, 오라클 PL/SQL 호환성을 개선한 전환도구인 마이그레이션 센터 6.11를 제공한다. 뿐만 아니라 단순 질의 최적화 기법으로 높은 성능을 제공하는 Altibase HDB DA 6.5도 제공한다.

큐브리드는 2014년 6월 Cubrid 9.3를 출시하였다. 이전 버전 대비 향상된 성능 및 기능을 제공하며 국산 오픈소스 DBMS로 포지셔닝하고 있다. 특히 클라우드 시장에 집중하며 국내 공개 소프트웨어 대표 기업들과 클라우드 통합 플랫폼인 'OCP(Open Cloud Platform)'을 출시하였다. 뿐만 아니라 한국공간정보통신의 인트라맵과 연동하며 지리정보시스템 시장에도 진출하고 있다.

리얼타임테크는 기존의 인메모리 DBMS 공간연산처리를 특화시킨 '카이로스 RDBMS'를 통해 시공간 DBMS로 포지셔닝 하고 있다. 또한 인메모리 DBMS를 활용한 실시간 합동표적관리 소프

트웨어 융합 제품 'RTeagle'를 출시하며 국방 분야 함정 전투 체계에 DBMS를 활용, 주요 표적에 대한 식별·분석 및 실시간 처리(융합, 비교일치, 위협평가 등)를 지원하며 분야를 확대해 나가고 있다.

선재소프트는 2013년 인메모리 기반 DBMS인 SunDB를 출시하며 시장에 소개되었으며, 현재 SunDB는 2.3버전이 최신 버전이다. 애플리케이션에 직접 접근하는 직접 액세스 방식을 통해 선형적인 확장이 가능한 것이 특징이다.

3. 기술 동향

최근 새로운 수요를 창출하기 위해 데이터를 활용하거나, 보다 많은 데이터로부터 통찰력을 얻기 위한 다양한 시도들이 이어지면서 기존 IT산업 환경에 변화가 불고 있다. 특히 데이터가 기존과 달리 크기(Volume), 속도(Velocity), 다양성(Variety)이 생기면서 데이터 저장부터 처리 및 관리에 대한 요구가 변화되고, 이를 구축·운영하는 IT 조직의 역할 자체에 변화가 일어나고 있다.

이 같은 변화를 수용하기 위해 IT는 효율성을 중요시하는 구조로 변모하고 있으며, 이 중심에 클라우드가 매개체가 되고 있다. DBMS 역시 기존 온프레미스 기반에서 장기적으로 클라우드 기반으로 재편될 것으로 보이며, 온프레미스 기반에서 운영되던 시스템과 새롭게 구축되어지는 클라우드 시스템 간의 효과적인 운영이 새로운 시대 IT 조직의 고민이 될 것으로 보인다.

향후 DBMS 분야에서 주목받을 기술로는 인메모리 DBMS, NoSQL DBMS, DBaaS(Database as a Service), DBPaaS(Database Platform as a Service), DBSaaS(Database Software as a Service), 오픈소스 DBMS, MPP(Massive Parallel Processing), Column-Store DBMS, DB 어플라이언스, 인공지능(AI), 딥러닝(Deep Learning) 등이 있다.

위에 언급한 기술들을 살펴보면 빅데이터와 클라우드가 중심이 되고 있으며, 보다 빠른 분석 및 실시간 요구를 수용하기 위한 기술들이 주목 받고 있다. 특히 클라우드 환경 하에서 데이터베이스 서비스에 대한 관심이 높아지고 있다. 단순한 사용량 기반 데이터베이스 서비스부터 플랫폼으로 제공되는 서비스까지 클라우드 상에서 데이터베이스 서비스는 매우 중요한 경쟁 요소로 자리잡을 전망이다.

이미 IT환경은 클라이언트-서버 기반의 레거시 환경에서 클라우드 환경으로 변화하기 시작하였으며, 물리적인 자원을 효율적으로 활용하기 위해 클라우드 인프라서비스(IaaS)가 매우 빠르게 확산되고 있다. 또한 데이터를 생성하는 애플리케이션 작업장이 기존 PC 중심에서 스마트기기 및 모바일로 확산되고 있다. 이는 데이터 생성 패턴 및 처리 과정의 급격한 변화를 불러일으키고 있다. 새로운 유형의 데이터를 활용한 비즈니스 영역이 생겨나고 있으며, 이 비즈니스들을 위해 기존 애플리케이션에서 발생하는 쌍방향 소셜 데이터들을 기반으로 한 데이터 분석, 실시간 분석에 대

한 요구도 데이터 관리 기술에 영향을 미치기 시작하였다.

기업 내 IT인프라에 대한 접근 방식도 변하고 있다. 기존 시스템 구매 방식에서 서비스 이용 방식으로 구매 및 소비 형태가 바뀌기 시작하였다. 이에 따라 각 기업들은 IT 자원을 효율적으로 활용하기 위한 방안으로 용도에 따라 클라우드 인프라 자원 활용을 고민하고 있으며, 이는 클라우드 환경을 지원하는 DBMS를 포함한 플랫폼까지 확대해 나가려는 움직임이 있다.

빅데이터 처리를 위한 기술 제공업체들의 다양한 시도들이 확대되고 있는데, 대표적인 것이 대량 데이터의 빠른 처리를 위한 인메모리 기술 활용, 많은 데이터를 처리하기 위한 MPP 아키텍처 도입, 기존 DBMS 성능을 개선하고자 하드웨어와 결합된 DB 어플라이언스, 적은 비용으로 대량 데이터들을 관리하기 위한 하둡 활용 등이다.

빅데이터 분석 기술에 대한 관심은 비정형 데이터뿐만 아니라 정형 데이터로 확대되고 있다. 특히 장기간 의무적으로 데이터 저장이 법제화 되는 경우나 금융권을 중심으로 주목받고 있는 이상거래 탐지시스템(FDS, Fraud Detection System) 등에서 보다 폭넓게 빅데이터 분석 기술들이 확산되고 있으며, 최근 인공지능 및 딥러닝 기술과 맞물려 더욱 주목받고 있다.

4. 시장 동향

국내 관계형 DBMS 시장이 라이선스 신규 매출보다는 유지보수 및 기존 업그레이드 등의 비중이 높아지고 있는 가운데, 데이터웨어하우스 교체, 금융권에서 계획중인 차세대 프로젝트, 기업데이터 증가로 인한 RDBMS 증설 등으로 관련 수요가 늘어날 것으로 전망된다.

국내 관계형 DBMS 시장은 글로벌 벤더인 오라클, 마이크로소프트, IBM 3개 회사의 매출이 전체 시장의 약 90%의 점유율을 보이고 있으며, 이 중 오라클은 60%가 넘는 점유율을 보이고 있다.

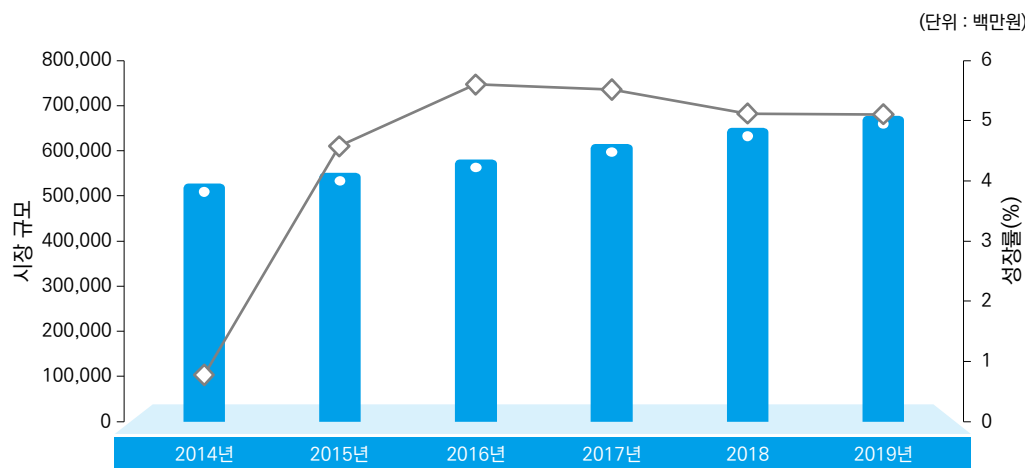
국내 벤더인 티맥스데이터, 알티베이스, 큐브리드, 선재소프트 등은 핵심업무용 DBMS 및 특성화된 DBMS를 통해 적용 사례가 빠르게 늘어나고 있다. 각사 제품의 강점을 살려 글로벌 제품의 전환 확대, 대형 볼륨 라이선스 계약 체결 등으로 비약적인 성장이 예상된다.

국산 DBMS 제품들의 제품 성숙도 확대, 고객 요구사항 수용, 국내외 기업과 협력을 통한 해외 진출, 공공기관과 기업의 국산 제품 도입 확산 등에 따라 2016년 국산 DBMS 점유율이 10% 이상으로 예측되며, 최근 몇 년간 꾸준히 점유율 상승에 따른 국산 벤더의 영향력 강화로 다양한 하드웨어 및 솔루션 업체들과 관계도 활성화되고 있다.

이외에도 빅데이터 제품과 서비스 확대, 클라우드 DB 서비스 제공 확산, 상용 오픈소스 DBMS 관심 증가, DW 어플라이언스의 지속 판매, x86확대에 따른 UNIX 플랫폼 감소 추세가 이어지고 있다.



그림 3-2-1 국내 RDBMS 시장 전망, 2015~2019년



※ 출처: IDC, 2015

5. 향후 전망

국내 관계형 DBMS 시장은 2015년부터 2019년까지 연평균 5.2%로 증가해 2019년 국내 관계형 DBMS 시장은 6,830억원 규모로 성장할 것으로 전망된다. 전체적으로 산업경기가 매우 저조한 타 산업군에 비해 견조한 상승세가 이어질 것으로 본다.

높아지는 IT 효율화 요구는 기업 업무 환경이나 IT 환경을 보다 변화시킬 것이다. 특히 스마트 기기의 확대에 의한 새로운 비즈니스 창출, 다양한 소셜 서비스에서 새로운 가치를 발굴하고자 하는 시도, 모바일이나 신규 OLTP 업무 확대에 의한 시장 창출, 분석 고도화를 위한 DW 고도화가 시장 성장을 이끌 것으로 전망된다.

빅데이터 분석 및 처리를 위한 다양한 기술적인 시도 역시 확대될 것으로 보인다. 새로운 유형의 데이터를 효율적으로 관리하고 기업 보유 데이터와의 통합 분석을 통한 새로운 뷰를 기반으로 기존의 DW를 확장하거나 새로운 BI를 접목하여 효과적인 의사결정을 위한 시각화 강화 등의 다양한 시도가 산업별로 확대될 것이다.

DB 어플라이언스의 경우도 하드웨어 노후화에 따른 틈새공략을 통해 기존 자사 제품을 사용하는 사용자들을 중심으로 하드웨어 고도화를 DB 어플라이언스로 해결하고자 하는 노력들이 심화될 것으로 보인다. 특히 대량 데이터 저장, 처리 및 분석 고도화, DB 통합(Consolidation) 등 고객 운영 상황에 특화된 전략들이 확대될 것으로 예상된다. 최근 국내에서 사례를 확보하고 있는 국산 DB 어플라이언스 제품과 외산 간의 경쟁이 더욱 치열해질 것이다. 향후 시장은 클라우드 시장의 주도권을 누가 갖느냐에 따라 급격한 시장 변화가 일어날 것으로 예상된다.

제4절 데이터 관리

1. 개요

조직의 의사결정에서 데이터의 중요성은 아무리 강조해도 지나치지 않다. 더구나 그 동안에는 조직 내에서 조직의 활동과 관련되어 발생한 데이터만 조직의 의사결정을 위한 데이터로 사용되었으나 이제는 조직을 둘러싼 외부 데이터, 더 나아가 조직과 직접적으로는 관계가 없던 데이터들까지도 조직의 의사결정에 필수 데이터가 되고 있어, 데이터 관리의 중요성은 조직의 내·외부 데이터의 구분이 없게 되었다.

그 동안 조직 내의 데이터는 주로 정형화 되어 관리되었던 관계로 이를 정형 데이터라고 하였다. 이에 비하여 조직 외부의 데이터는 그 형태를 ‘정형’이라는 틀에 담기에는 기술적 한계가 있어서 이를 비정형 데이터라고 하고, 데이터의 규모가 테라바이트를 넘어 페타바이트 규모에까지 이르러 이를 빅데이터라고도 한다.

정형 데이터의 관리에서 데이터를 보관하고 이를 응용 프로그램들이 읽거나 사용하는 전통적인 OLTP에서는 주로 관계형 DBMS가 사용되었고, 데이터 관리의 중요성이 높아지면서 이를 위한 많은 솔루션이 시장에 등장하고 있다. 그 동안 데이터 관리 관련 솔루션 시장은 주로 외산 솔루션에 의해서 주도되었지만, 최근 국내 소프트웨어 개발사들이 외산 못지 않은 좋은 솔루션을 제공하기 시작하여 시장에서 국산 솔루션의 점유율이 점차 높아가고 있는 추세이다.

최근 사회 전반으로 정보시스템에 대한 의존도가 높아지고 정보시스템에 대한 요구사항도 많아지고 복잡하게 되었다. 따라서 이에 대응하기 위한 데이터의 용량도 자연스레 커지게 되었다. 데이터의 관리도 기존의 관리 범주에서 데이터 성능관리와 데이터 품질관리, 데이터 보안관리 등으로 그 범주가 커지면서 많은 관련 솔루션들이 시장에 나오게 되었다.

반면 비정형 데이터의 경우는 아파치 라이선스의 하둡을 기반으로 솔루션이 발전하고 있어 솔루션 자체로는 시장이 크게 형성되지 못하고 그 대신 하둡을 기반으로 한 전문가 컨설팅 시장이 형성될 가능성이 높을 것으로 예상된다.

2. 제품 동향

정형 데이터의 전통적인 데이터 관리 솔루션인 DBMS는 주로 외산 제품들이 시장을 장악하고 있다. 최근 들어 티맥스와 알티베이스, 큐브리드 등 국산 솔루션들이 등장하면서 국내 시장에서 조금씩 그 사용 정도가 높아가고 있지만, 아직도 오라클, 사이베이스 등을 비롯한 외산 DBMS가 데이터 관련 솔루션 시장의 거의 대부분을 장악하고 있다. 데이터 설계의 경우도 ERWin이나

ER/Studio 등 외산 제품이 시장을 거의 장악하고 있는 가운데 국내 기업인 엔코아의 DA# 등이 조금씩 그 시장을 넓혀 가고 있다.

비정형 데이터의 경우 데이터 관리 솔루션은 하둡을 기반으로 하둡의 생태계를 넓혀가거나 혹은 하둡의 한계를 개선한 제품이 나오고 있는데, 국내의 경우 그루터가 타조(Tajo)라는 솔루션과 이를 중심으로 데이터를 수집·저장·분석하는 TaaS(Tajo as a Service) 솔루션, 기업용 데이터 처리용 Tajo Enterprise Geer 솔루션을 시장에 출시하고 있다. 그 외에도 다음소프트, 데이터스트림즈 등에서 빅데이터의 수집에서부터 처리, 분석, 관리 등과 관련한 많은 솔루션을 출시하고 있다.

3. 기술 동향

정형 데이터의 관리 기술은 우선 데이터를 물리적으로 적재함에 따른 효율적 알고리즘과 관련된 기술의 발전이 주류를 이루었으나, 최근에는 데이터 관리와 관련된 사용자 편의성 증진을 위한 여러 기능의 발전이 주류를 이루고 있다. 또 데이터 품질관리 관련 기술의 경우에는 데이터의 정확성을 위하여 업무규칙이 세분화되고 전문화된 방향으로 발전하고 있는데, 향후에는 AI를 사용하여 데이터의 정확성, 일관성 등 데이터 품질과 관련된 진단을 하는 방향으로 발전될 가능성이 높다.

한편 정형 데이터에서 성능관리를 위한 기술은 성능이 DBMS의 내부 알고리즘과 밀접한 관련이 있어 이에 대한 연구가 계속되고 성과 또한 지속적으로 나타나고 있다. 최근에는 인메모리 기술을 DBMS에 적용하여 데이터를 디스크에서 읽어 오던 기존 방식에 비해 혁신적인 성능 향상을 가져오고 SSD(Solid State Drive)가 서버의 HDD(Hard Disk Drive)를 대체함에 따른 성능 향상도 기대된다. 데이터 보안의 경우에는 주로 데이터의 암호화에 따른 성능 부하를 줄이기 위한 방향으로의 암호화 알고리즘 기술 개선이 주요 흐름을 이루고 있으며, ETL(Extraction, Transformation, Loading) 등 데이터 흐름과 관련된 기술에서는 대용량 데이터의 빠른 적재를 위한 기술이 SSD의 도입과 함께 발전의 주요 흐름을 이룰 것으로 예상된다.

비정형 데이터의 관리 기술은 구글 GFS(Google File System)와 맵리듀스(MapReduce)기술에서 파생된 오픈 소스인 하둡 생태계에 기반을 두고 관련 기술이 발달되고 있다. 플룸(Flume)은 대량 로그 데이터를 효율적으로 수집하고 장애 발생 시에도 데이터를 잃지 않고 페일오버(Failover) 기능을 갖고 있으며, 여러 대의 에이전트(Agent)나 콜렉터(Collector)와 연동이 가능하다. 데이터 저장 솔루션으로는 HDFS(Hadoop Distributed File System)가 비용 효율적인 x86PC 사양의 노드를 스케일 아웃 아키텍처에 의한 분산 파일시스템을 제공하고 있고, 맵리듀스가 HDFS에 저장된 빅데이터를 용이하게 분석할 수 있게 하여 이들이 해당 솔루션의 중심 기술로 자리하고 있다. 현재 국내의 많은 기업이 빅데이터의 솔루션을 개발하고 있는데 이들 대다수가 하둡을 기반으로 하고 있다.

4. 시장 동향

데이터 성능 관리 솔루션은 2010년 15%를 넘는 고성장에서 현재는 다소 낮은 5%대의 성장을 보이고 있다. 성능관리 시장은 DBMS 시장에서 오라클의 둔화와 인메모리 기술이라는 컴퓨터 발전에 대한 데이터 성능 의존도가 높아짐에 따라 데이터 성능과 관련한 시장 성장이 다소 완화되는 경향이 있다. 반면, 보안 관련 데이터 관리 솔루션은 2013년까지 매년 거의 30%에서 60%에 이르기까지 초고도 성장을 하였으나 최근에는 10% 밑의 성장을 보이고 있다. 이는 지난 2011년, 2012년, 2013년 등 3년 동안의 보안에 대한 집중적 투자에 대한 반작용으로 향후에는 다시 고도 성장이 가능하리라 예상된다.

비정형 데이터 솔루션 시장은 주요 기반 솔루션인 하둡이 아파치 라이선스의 오픈 소프트웨어라 정형 데이터 솔루션 시장에서 형성되었던 고도의 성장이나 대형 시장의 형성을 기대하기는 어려울 것으로 전망된다. 대신 빅데이터 솔루션을 도구로 한 컨설팅 시장의 성장 가능성은 클 것으로 예상된다.

5. 향후 전망

데이터 모델링 툴은 데이터 구축을 위한 여러 툴들과의 통합이 가능한 형태로 발전할 것으로 전망된다. 즉 데이터 설계에서부터 구축, 운영, 검증 등의 작업이 통합된 툴에서 끊임없이 수행될 수 있도록 발전될 것이며, 향후 설계 툴이 이런 형태로 발전되었을 때 정형화 데이터의 솔루션 시장을 주도하는 솔루션으로 성장할 것으로 전망된다.

DBMS는 전통적인 OLTP 및 DW의 중심 솔루션으로 계속 발전할 것으로 전망되며 인메모리 기술의 발전과 SSD 등 컴퓨터의 물리적 발전에 따른 하드웨어 가격의 지속적인 하락 등으로 성능과 관련된 기술은 DBMS 보다는 컴퓨터 기술에서 치열하게 경쟁하리라 예상된다.

데이터 품질과 관련된 솔루션은 데이터 거버넌스 솔루션으로서 진화를 할 것이고 현재 개별적으로 작동하는 데이터 품질과 관련된 모든 솔루션들이 데이터 거버넌스 솔루션으로 통합될 것이다. 향후에는 모든 데이터 관리 솔루션과 데이터 품질 관련 솔루션이 데이터 거버넌스 솔루션으로 통합되리라 예상된다. 데이터 보안과 관련된 솔루션의 경우는 DB 암호화와 접근통제를 중심으로 계속 발전하고, 개인정보 보호법과 관련된 솔루션의 등장이 예상된다. 또 클라우드 서비스 환경이 보다 일반화되어 감에 따라 인하우스에서 갖고 있던 보안 약점들을 보완하는 기술과 솔루션이 틈새시장용 솔루션으로 나오리라 예상된다. 현재 거의 대부분 솔루션들이 오픈 소스를 지향하고 있지만, 보안 솔루션이 야말로 공개를 통한 보안 약점 개선에 대한 기대 때문에 오픈 소스가 주된 흐름이 되리라 전망된다.

비정형 데이터 및 빅데이터 솔루션의 경우는 현재 아파치 라이선스의 하둡을 기반으로 하고 있어서 대부분의 솔루션이 솔루션 자체만으로 시장을 형성하기보다는 빅데이터 전문가의 컨설팅 서비스에 사용되는 도구로 그 역할을 할 것으로 예상된다. 향후 빅데이터와 관련하여 가장 기술적으로 큰 발전이 기대되는 부분은 사용 용이성 증대와 데이터 시각화의 발전이라고 할 수 있다.

제5절 데이터 보안

1. 개요

전통적인 웹 보안, 데이터베이스 보안, 파일 보안, 네트워크 보안은 가상화, 클라우드 환경, 빅데이터, IoT에 대한 생태계가 발전하면서 크게 달라지고 있다. 가트너에 따르면 2016년 60억개의 IoT 디바이스가 연결되고 2020년이 되면 200억개 이상의 IoT 디바이스가 연결될 것으로 예측하지만, 각종 보안 취약점으로 인한 데이터 유출 사고가 더욱 증가할 것으로 언급하였다.

각 분야의 수많은 보안 기술 기업들과 연구기관들이 새로운 보안기술을 내놓고 있지만, 이를 우회하기 위한 해커들의 기술 발전도 명확한 사실이다. 「2016 데이터 유출 보고서(ITRC Data Breach Report, Jun 2016)」에 따르면, 2015년에 보안 부문 투자가 전년 대비 22% 증가했음에도 불구하고, 데이터 유출 사고가 38% 증가했다. 데이터 유출 사고에 대한 고민은 공공기관 및 일반 기업들의 담당자 혹은 CIO들에게 여전히 가장 중요한 이슈 중 하나이며, 빅데이터 분석과 클라우드 쪽으로의 환경 변화로 인한 보안 이슈는 더욱 커지게 되었다. 또한, 국내 개인정보 보호법에 따른 정보보호 관리 체계(ISMS, Information Security Management System) 인증, 개인정보 영향평가 등과 같은 법적 대응 이슈, 의료정보 보호법(HIPPA, Health Insurance Portability and Accountability), PCI-DSS(Payment Card Industry Data Security Standard)와 같은 각종 산업별 IT 컴플라이언스에 대응하기 위한 기술적 조치가 증가하게 되었고 이러한 현상은 한동안 지속될 것이다. 데이터보안은 IT 자산 중에서 가장 핵심이 되는 데이터베이스 보안에 이르게 되며, 이는 모든 IT 인프라에 제공되는 데이터의 양과 기초는 데이터베이스에서 출발하기 때문이다. 데이터베이스 보안은 데이터의 기밀성, 무결성, 가용성이라는 목표를 달성하기 위해 DB접근제어, DB 암호화, DB 작업결재, DB 취약점 분석으로 나뉘며, 이는 한국데이터진흥원에서 제시하는 데이터베이스 보안 가이드라인 및 프레임워크(DQC-S)에서도 이에 대한 기술적/관리적 조치 가이드를 제시하고 있다. 데이터베이스 보안을 위해서는 단순한 솔루션 도입이 아니라 기획·설계·구축·운영 단계별로 준비하는 것을 요구하고 있다.

2. 제품 동향

DB 접근제어 솔루션의 경우, 외산제품은 포티넷(Fortinet)의 FortiDB, IBM의 가디엄(Guardium), 임퍼바(Imperva)의 SecureSphere, 맥아피(McAfee)의 Data Centric Security Suite, 트러스트웨이브(Trustwave)의 DbProtect가 대표적이며, 국산제품으로는 웨어벨리의 Chakra MAX, 피앤피시큐어의 DB Safer, 신시웨이의 Petra, 소만사의 DB-i가 대표적이다.

국산 DB 접근제어 솔루션들은 DB 작업결재 기능이나 시스템 접근제어 기능을 동시에 가진 제품들로 인해 국내 고객의 요구사항을 충족시키고 있으며, 더 나아가 DB 암호화 기능과 연계되는 기능을 제공하기도 한다. 이에 반해 외산제품들은 DB 접근제어와 함께 취약점 분석 기능을 연계하는 경우가 많다.

DB 암호화 솔루션의 경우, 외산제품은 HP Security Voltage, Oracle Advanced Security, Protegrity Data Security Platform, Vormetric Transparent Encryption가 대표적이며, 국산제품으로는 케이사인의 Secure DB, 펜타시큐리티의 D'amo, 이글로벌시스템의 Cubeone, 웨어벨리의 Galea, 소프트포럼의 XecureDB가 대표적이다. DB 암호화 솔루션은 칼럼 레벨의 암호화 방식과 파일 레벨의 암호화 방식으로 나눌 수 있는데, 파일 레벨의 경우 암호화로 인한 속도 저하가 적지만 사용자 식별이 어렵다는 단점을 가지고 있다. 이는 최근 ISMS 인증평가나 DQC-S 인증평가에서도 지적되고 있다. 칼럼 레벨의 암호화 경우, 플러그인 타입과 API 방식이 존재하고 업무의 성격에 따라 혼합하여 사용하는 경우가 많다. DB 접근제어 구성과 유사한 프록시 방식의 DB 암호화를 제공하는 제품과 토큰화(Tokenization) 방식의 암호화 제품도 있으나, 국내에서 적용사례가 칼럼 암호화 방식에 비해 많지 않다.

국내 시장에서 DB 암호화 제품의 경우, IT보안인증사무국에서 발표하는 'DB 암호제품 보안 요구사항' 규격을 만족하여야 하며, 암호검증모듈탑재 및 최근 CC인증으로 유도되고 있다.

DB 취약점 분석 솔루션의 경우, 외산 제품으로는 AppDetective, Rapid 7이 대표적이며, 국산 제품으로는 웨어벨리의 Cyclone, 나일소프트의 Secuguard가 대표적이다. 최근 DB 취약점 솔루션은 데이터 보안진단 솔루션으로 발전하여, 개인정보 스캔, 데이터 보안 현황 분석, 포렌식 분석 등 다양한 기능으로 개인정보 영향평가나 ISMS 인증 평가에 대응하고 있다.

최근 빅데이터 시장의 성장에 따라, NoSQL 기반의 빅데이터 플랫폼을 기존 RDBMS와 연계하는 사례가 증가하면서, 이로 인한 보안 이슈가 동시에 발생하고 국내외 전문 기업들이 관련 솔루션을 제공하고 있다. 하지만 여전히 빅데이터에 대한 데이터 보안 기술이 성숙하지 못하고, 빅데이터에서의 프라이버시 문제와 함께 고객 요구에 명확히 충족되는 제품이 존재하지 않는다.

3. 기술 동향

가. DB 접근 패턴 학습 기능(Machine Learning)/ 이상 징후 차단

IBT(International Business Times)에 따르면, 유아용 전자기기 제품을 제공하는 VTech은 자사의 데이터베이스 보안 소홀로 인해 2015년 8,000만명의 고객정보가 침해당했다. 이외에도 '2015 ITRC Data Breach Report'에 따르면 2015년 781건의 데이터 유출사고가 발생했으며, 이로 인해

1,700억건의 개인정보(사회보장번호, 운전면허번호, 의료 및 금융기록)가 유출되었다. 전문가들은 이러한 침해사고는 DB접근제어(DAP, Database Audit and Protection)에서의 데이터베이스에 대한 접근 패턴을 학습하고 이상 징후 시 이를 차단하는 기술이 필요하다고 언급한다. 또한 대부분의 데이터 유출 사고는 보안 솔루션을 우회하거나 인가된 사용자를 위장한 공개된 DB 취약점을 통한 공격에 의한 것이다. 이것이 오라클, SAP, 마이크로소프트, IBM 등 수많은 DBMS 벤더들이 자사의 취약점을 정기적 혹은 비정기적으로 공개하고 패치를 유도하는 이유이다.

나. 가상화/클라우드 공간에서의 DB보안 기술

국내의 경우 2016년 말까지 보안 강화를 위해 금융권 망 분리를 구축해야 하고, 이로 인해 다양한 가상화 환경과 클라우드 환경이 기존 인프라에 연계되고 있다. 수많은 기업내 IDC에 존재하던 데이터베이스들이 가상화/클라우드 공간으로 이동하면서 데이터베이스 보안에 대한 기술도 진화하고 있다. 아마존, IBM, 오라클, KT 등의 대표적인 클라우드 환경 그리고 EMC, 레드햇, 마이크로소프트, 시트릭스 등의 가상화 솔루션 환경이 서로 교차되면서 고객들의 환경도 다양하게 변했으며, 이로 인한 DB 보안 기술의 환경 적응 사례가 나타나고 있다. 국내외 DB 접근제어, 암호화, 작업결재, 취약점 분석 솔루션들은 이미 이러한 환경에 대비한 솔루션들을 제공하기 시작했다.

다. DB 암호화-DB접근제어의 연동

DB 접근제어를 경유한 DB 접속 사용자는 DB에 설치된 암호화 솔루션에서 사용자 식별(End-User IP)이 어려운 경우가 자주 발생한다. 이로 인해 복호화 권한 부여에 어려움이 증가하게 되어, DB 접근제어와 연동하여 사용자 식별을 구분해 낸다.

라. 사용자 단말에서의 개인정보 유출 통제

인가된 DB 사용자(DBA, 개발자 등)가 업무의 필요에 따라 DB 내의 개인정보나 기업의 민감정보를 조회하는 경우가 다수 발생하게 되며, 이메일, USB 등과 같은 매체를 통해 데이터 유출 위험이 발생된다. 기존의 DLP 솔루션에 의해 이를 통제하는 경우도 존재하지만, 최근 DB 보안 솔루션에서 이를 제어하는 기능도 고객들에 널리 보급되고 있다.

4. 시장 동향

기업의 데이터베이스에 대한 보안 프로세스 강화에도 불구하고 APT, 랜섬웨어와 같은 지능적인 공격 증가에 대비하려면 미들웨어, 응용프로그램, 데이터베이스에 대한 꾸준한 보안패치, 다양한

데이터 보안의 기술적 조치와 모니터링만이 데이터 유출사고를 막는 최적의 방안이다. 포레스터(Forrester) 보고서에 따르면, 2014년부터 데이터베이스 보안 시장은 매년 20%씩 꾸준히 성장할 것이라 전망했다. IBM, 마이크로소프트, 오라클, SAP 등의 DBMS 벤더들뿐만 아니라, 임퍼바, 트러스트웨이브, 포트넷 등 독립 소프트웨어 벤더들도 지속적으로 이 시장에 파고들고 있다. 데이터베이스 보안 영역은 보안정보 이벤트관리(SIEM, Security information and event management), 침입방지시스템(IDP, Intrusion detection and prevention), 데이터 손실방지(DLP, Data leak prevention)와 같은 다른 보안 영역과 통합되는 움직임도 빨라지고 있다. 국내 시장은 CC인증 요구사항 및 커스터마이징, 그리고 빠른 대응 필요에 따라 국산 데이터베이스 보안 업체들이 시장을 장악하고 있다.

데이터베이스 보안시장의 빠른 성장으로 인해 이미 5~6년 전에 IBM은 가디엄을 인수하였으며, 오라클은 서체르노(Secerno)를, 맥아피는 센트리고(Sentrigo)를 인수하였다. 2015년에는 HP가 볼티지 시큐리티(Voltage Security)를 인수, 싱텔(Singtel)은 트러스트웨이브를 인수, 탈레스(Thales)는 보메트리(Vormetric)를 인수함으로써 글로벌 데이터베이스 보안 시장은 점점 치열해지고 있다. 이미 클라우드와 가상화 시장이 성숙되고, IoT 및 빅데이터 활성화로 인해 각 벤더들은 데이터베이스 보안을 위해 각종 기술들을 시장에 출시하고 있다.

5. 향후 전망

IDC에 따르면 데이터베이스 시장은 현재 400억달러 규모이고, 2017년에는 500억달러에 이를 만큼 거대하고 안정적인 성장 시장이다. 보안 시장에 대한 기업들의 투자 중 40%는 기존 방화벽, IDS와 같은 네트워크 보안에 투자되고 있는 반면, 데이터베이스 보안에 대한 투자는 19%에 머무르고 있다. 그러나 사이버 범죄가 꾸준히 증가함에 따라, 데이터베이스에 대한 보안 투자도 증가 추세를 보인다. 수많은 글로벌 기업들이 데이터 보안 솔루션의 인수합병을 통해 고객들의 투자 흐름에 대비하고 있다.

제6절 데이터 품질

1. 개요

데이터는 기업의 가장 큰 자산으로서 데이터 활용을 기반으로 하는 마케팅 전략 수립, 지표데이터를 기반으로 한 위험 관리 및 의사결정 등을 통한 경영 활동은 이미 오래전부터 그 노력이 계속되고 있다. 또한 최근 뜨거운 이슈로 떠오르고 있는 인공지능, 기계학습 등을 통한 빅데이터 서비스 분야도 점차 늘어나고 있다. 이에 데이터가 생산·취합되는 위치에서의 데이터 품질이 그 어느 때보다 절실히 필요한 시점이다.

데이터 품질관리 솔루션은 데이터의 일관성, 정확성 등을 확보하기 위해서 데이터 품질 활동에 대한 프로세스 및 기능을 제공하는 솔루션으로 초기 금융권 구축을 시작으로 급속히 확산되어 근래에는 전 산업 분야로 영역을 넓혀가고 있다.

2. 제품 동향

국내 시장은 데이터 품질관리 체계 수립에 따른 조직 및 프로세스 등과 관련한 커스터마이징 요구사항들이 발생하다보니 국산 솔루션들이 시장을 주도하고 있다. 국내 데이터 품질관리 솔루션 기업으로는 위세아이텍, 지티원, 데이터스트림즈, 엔코아 등이 있으며, 솔루션 외에도 자체적인 데이터 품질관리 방법론을 보유하고 있다.

위세아이텍의 'WISE DQ'는 웹 애플리케이션으로 배포되어 설치 시 OS에 종속적이지 않으며 HTML5 기반으로 구성되어 있어 멀티브라우저를 지원하는 등 접근성과 확장성, 이식성이 우수하다. 주요 기능으로는 자동화된 데이터 프로파일링, 데이터 오류 감시, 데이터 규칙 관리, 데이터 품질 분석 및 분석 결과에 대한 다양한 통계 보고 등을 제공한다. 데이터 오류의 원천을 신속히 파악할 수 있도록 데이터 매핑 정의서 기반의 데이터 흐름 정보를 다이어그램으로 제공하기도 하며, 자사 메타데이터 관리 시스템인 'WISE META'와 연계를 통해 데이터 표준 기반의 데이터 품질관리가 가능하다.

표 3-2-1 데이터 품질관리 솔루션 주요 기업 및 제품

기업명		제품명
국내 기업	데이터스트림즈	Quality Stream
	엔코아	DataWare DQ
	위세아이텍	WISE DQ
	지티원	DQMiner
글로벌기업	IBM	InfoSphere QualityStage
	Informatica	Data Quality
	SAS	DataFlux

※ 출처: 각사 자료 취합

지티원의 ‘DQ Miner’는 자동 데이터 프로파일링, 데이터 오디팅, 데이터 룰 관리, 데이터 품질 분석, 데이터 품질 분석 결과 보고 등을 제공하는 전사 데이터 품질관리 솔루션이다. 이것은 고유의 데이터 품질관리 방법론을 제공하여 기업이 데이터 품질 목표를 달성할 수 있도록 상시 데이터 품질 모니터링 시스템을 구축할 수 있게 해준다. 또한 자사 애플리케이션 영향도 분석 솔루션과 연계하여 오류 데이터의 원인을 쉽게 파악할 수 있으며 애플리케이션 추적 관리, 오류 데이터 정제 관리가 가능하고 80여 종의 데이터 품질 관련 보고서를 제공한다.

데이터스트림즈의 ‘QualityStream’은 분석대상 데이터에 접근하여 품질 진단, 결과 도출 및 분석을 하여 지속적인 데이터 품질관리의 유지 및 향상을 확보하기 위한 솔루션으로 분석대상 데이터베이스를 프로파일링 수행하여 현재 상태의 품질 수준을 분석 후 지속적인 관리 대상을 등록 및 스케줄링하여 결과를 분석하고 오류사항에 대한 정비 프로세스를 활용하여 관리한다. 통합 레파지토리 구성을 통한 프로파일 관리 기능, 비즈니스를 관리 기능, 품질 진단에 대한 특정 기준별 결과 검색과 통계 정보를 제공하는 검증 결과 관리 기능, 오류 데이터에 대한 정비 프로세스를 지원하는 정비 관리 기능을 제공하며 자사 CDC 솔루션과의 연계를 통한 품질관리를 지원한다.

엔코아는 기업의 목표 달성을 위한 비즈니스 환경 예측과 성과 지향적 데이터 관리 체계 수립을 위한 설계부터 관리에 대한 토털 데이터 솔루션을 단일 레파지토리 기반으로 제공하고 있다. 이 중 ‘DataWare DQ’는 자사 모델링 도구인 ‘DataWare DA’, 메타데이터 관리 솔루션인 ‘DataWare META’와 연계하여 데이터 설계를 기반으로 한 데이터 품질관리를 제공한다. 주요 기능으로 데이터 품질관리 기준 정의 기능, 프로파일 및 업무규칙 관리 기능, 품질진단 실행 기능, 측정 결과 분석 기능, 오류 원인 개선 관리 기능을 제공한다.

외산 솔루션으로는 IBM의 ‘Information Analyzer/QualityStage’, 인포매티카(Informatica)의 ‘Informatica DataQuality(IDQ)’, SAS의 ‘DataFlux’ 등이 있다.

IBM의 'InfoSphere QualityStage'는 주소 및 이메일, 전화번호, 부품명, 상품명 등 텍스트 형태의 데이터를 마스터데이터로 표준화할 수 있는 소프트웨어 솔루션이다. 기업들의 주소 표준화 요구에 따라 주소 표준화 기능을 주요 기능으로 제공한다. 마스터데이터 표준화 기능에 주소를 포함한 텍스트 기반의 데이터 제공이 가능하다.

인포매티카의 'Informatica Data Quality(IDQ)'는 논리적인 비즈니스 규칙을 구축할 수 있는 'Business Rule builder' 기능을 제공하여 현업 부서 사용자가 IT에 의존하지 않고 비즈니스 규칙을 신속하게 개발할 수 있도록 지원한다. 또, 수동 작업을 워크플로에 통합하여 현업 부서 사용자가 자동화된 프로세스 전체에서 예외를 검토, 수정 및 승인할 수 있도록 지원한다.

SAS의 'Data Quality'는 비표준적인 기록이나 중복 기록과 알려지지 않은 데이터 유형을 수정할 수 있는 데이터 정리, 테이블, 데이터베이스, 소스 애플리케이션에서 나타나는 관계들을 파악해 데이터를 보다 잘 이해할 수 있는 데이터 프로파일링 기능을 제공하여 통합된 웹 기반 콘솔을 통해 데이터 품질 작업을 모니터링하고 데이터 문제 및 거버넌스 활동을 확인할 수 있도록 지원한다.

3. 기술 동향

데이터 품질관리 솔루션은 데이터 품질을 확보하기 위해 데이터의 규칙을 도출하여 이에 위배되는 데이터를 진단하고 그 결과에 대해 원인 분석 및 개선 활동을 수행하는 등 품질개선 노력에 대한 모니터링 및 과정에 대한 프로세스 기능을 제공한다.

데이터 규칙을 도출하기 위해서는 데이터에 대한 정확한 이해가 선행되어야 하고, 간단하고 정형화된 규칙의 경우 데이터 정의로부터 보다 편리하게 도출할 수 있다. 최근에는 데이터 품질관리 솔루션과 모델링 도구, 애플리케이션 영향도 분석 도구, 메타데이터 관리 도구 등의 솔루션을 함께 도입하여 상호 연계를 통해 프로파일 자동 연계 등록, 메타데이터 등과 싱글뷰(single view) 구현, 변경관리 시점에서의 품질 점검 등을 수행한다. 또한 영향도 분석 솔루션과 연계를 통해 오류 데이터 원인 추적 등 데이터 품질관리를 위한 활용성을 극대화하려는 사례가 늘고 있다.

메타데이터 관리와 데이터 품질관리 분야는 각각 데이터의 사전적, 사후적 관점의 관리이지만, 데이터 관리 업무 관점에서 본다면 하나의 연속된 업무로 여겨질 수 있다. 때문에 메타데이터 관리와 데이터 품질관리 솔루션을 통합 구축하거나 통합 환경을 제공하기 위한 데이터 거버넌스 포털을 구현하는 사례도 많아지고 있다.

최근 공공데이터 개방과 관련하여 법안과 지침에 따른 데이터 품질 요소가 필수로 요구되기 때문에 시스템과의 연계를 통한 실용적인 데이터 품질관리 프로세스 적용이 필요할 것으로 보인다.

현재까지는 주로 정형 데이터에 대한 데이터 품질관리를 적용했으나, 빅데이터를 위한 비정형 데이터에 대한 데이터 품질관리가 실현될 수 있는 기술적 요소들이 추가되어야 할 것이다.

4. 시장 동향

데이터 품질관리 솔루션은 오류 데이터를 식별하고 정제할 목적으로 초기에 도입되었고 점차 표준, 구조, 흐름, 관리 체계 수립 등 데이터 품질과 관련된 여러 분야로 확대 발전해 왔다. 차세대 나 시스템 통합과 같은 정보화 사업의 일환으로 관리 체계가 구축되었으며, 특히 구 시스템으로부터 신 시스템으로의 데이터 이관 과정에서 데이터 정합성 확보를 위해 데이터 품질관리 솔루션이 많이 활용되었다.

근래에는 공공데이터 개방과 맞물려 개방 데이터에 대한 품질 이슈로 관련 솔루션이 다양하게 활용되고 있다. 시장 규모는 작지만 데이터의 중요성이 부각되고 있으며, 관련 기관과 기업들의 꾸준한 노력으로 매년 10% 내외의 성장세를 유지해 왔다.

향후에도 정부 차원의 데이터 품질관리 요구와 정확한 데이터 기반의 빅데이터 분석 요구가 맞물려 시장 전망은 매우 밝을 것으로 예상된다.

5. 향후 전망

2016년 초 알파고와 이세돌 기사의 바둑 대결은 인공지능의 현재 위상을 다시 한번 일깨워 준 이벤트였다.

인공지능, 기계학습, 딥러닝, 클라우드, 빅데이터, IoT 등 기술 용어들이 제4차 산업혁명에 진 입했음을 알리며 화려하게 입장하고 있다. 하지만 이러한 기술들도 고품질의 데이터를 기반으로 할 때 존재 가치를 부각시킬 수 있다. 기업 자산으로서의 데이터, 이를 위한 데이터 관리 노하우가 기업 경영을 위해 무엇보다 선행되어야 하며 데이터 관리 체계에 따른 양질의 데이터 생산이 향후 기업 성장에 중요한 요소로 작용할 수 있을 것이다.

공공부문에서는 공공정보 개방 및 공유 분야 시스템 구성 등에 데이터 품질관리 솔루션 활용이 점차 확대될 것으로 보인다. 데이터 개방·공유 이전에 해당 데이터의 품질을 사전 점검 후 개방·공유하는 프로세스와 기능 제공도 필요하다.

데이터의 폭발적 증가에 따라 데이터 품질관리 적용 분야는 앞으로 더욱 확대될 것으로 보인다. 시장의 성장과 함께 데이터 활용기술의 변화에 따라 데이터 품질관리 기술도 끊임없이 진화할 전망이다.

제7절 데이터 분석

1. 개요

데이터 분석은 다양한 과학적, 공학적 기법들을 통해 반복적으로 데이터를 탐색 및 조사하여 데이터 속에 숨겨진 패턴을 발견하고 전체를 올바르게 이해하여 인사이트(Insight)를 얻고 그 결과로 합리적인 의사결정과 비즈니스 성과를 높이는 것을 목적으로 한다. 데이터 분석은 크게 비즈니스 인텔리전스(BI, Business Intelligence)와 고급분석(AA, Advanced Analytics)으로 구분할 수 있다. BI 업체들은 AA 제품을 개발하기 위해 기존 제품에 셀프 서비스(Self-Service) 기능을 강화하고 있으며 머신러닝(Machine Learning)의 예측(Prediction) 기능 등을 강조하는 제품과 고급분석 기능을 제공하는 제품으로 발전하고 있다. 기술 측면에서는 스파크 기반으로 하둡 환경이 빠르게 재편되고 있고 오픈소스 R, 파이썬(Python) Pandas 기반의 분석 언어가 인기를 얻고 있다. 또한 구글 클라우드 플랫폼을 기반으로 한 머신러닝 서비스가 출시되는 등 BDaaS(Big Data as a Service) 서비스에 대한 관심이 늘어나고 있는 추세이다.

표 3-2-2 비즈니스 인텔리전스와 고급분석 비교

	비즈니스 인텔리전스	고급분석
지형	과거	미래
질문 유형	무엇이 발생했는지. 언제, 누가, 얼마나	왜 발생하는가, 이 트렌드가 계속 될 것인가? 다음에 무엇이 발생할 것인가?
방법론	보고서(KPIs, 지표) 자동화된 모니터링/알림 대시보드 OLAP(큐브) 임의(Ad-Hoc) 쿼리	예측 모델 / 탐색 모델 데이터마이닝 텍스트 마이닝 멀티미디어 마이닝 통계적 / 정량적 분석 시뮬레이션 / 최적화
빅데이터 여부	예	예
데이터 타입	구조화된 데이터	구조화 / 비구조화 데이터
지식 생성 방식	수동적	자동적
사용자	비즈니스 일반 사용자	데이터 과학자, 비즈니스 분석가, IT 관계자, 비즈니스 일반 사용자
의사결정방식	결과 기반 의사결정	선제적 의사결정

2. 제품 동향

데이터 분석을 위한 제품은 전통적인 BI 제품, 고급분석 중심의 AA 제품과 기존 BI 제품에 고급분석 기능을 추가한 하이브리드 제품으로 나눌 수 있다.

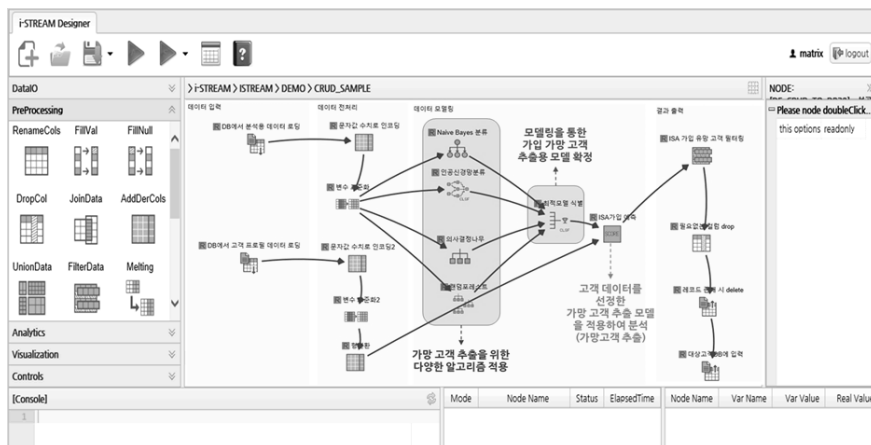
국내 기업으로는 비아이매트릭스의 MATRIX Suite, 야인소프트의 Octagon EOS, 위세아이텍의 WISE OLAP, 오브젠의 eCube 등이 있다. 이 제품들은 D3.js와 Bokeh 기반의 시각화 기능 확장, 오픈소스 R 기반의 분석 연계, 인메모리 분석, 하둡 연동 커넥터 등으로 확장하여 제품화하고 있다. 외산 솔루션인 SAP BO와 IBM Cognos 역시 다양한 기능을 강화하고 있지만 전통적인 OLAP 리포팅에 머물러 있고, 전문 시각화 제품으로 두각을 보이는 있는 클릭뷰(Qlikview), 스폿파이어(Spotfire), 태블로(Tableau)가 주목을 받고 있다.

고급분석을 위한 오픈소스 R 기반의 분석은 샤이니(Shiny), 지지플롯2(ggplot2)가 시각화 기능과 클라우드 연동 등으로 확산되고 있다. 파이썬 아나콘다(Python Anaconda) 배포판의 쉬운 설치, IPython-note 등의 편리한 개발 환경 때문에 Pandas 기반 분석 방법도 인기를 얻고 있다. 셀프서비스 기능을 제공하는 외산 제품으로는 SAP의 Lumira, SAS의 Visual Analytics 등이 있다.

클라우드 기반 분석으로는 마이크로소프트 애저(Microsoft Azure)의 Cortana Intelligence Suite와 구글 클라우드 플랫폼 등이 있으며, 이들 제품은 머신러닝 알고리즘을 이용하여 빅데이터 예측 분석을 제공하고 있다.

BI 솔루션에 고급분석 기능을 추가한 기능 확장도 활발하게 진행되고 있다. 비아이매트릭스와 마이크로스트래티지(MicroStrategy) 등은 머신러닝 기반의 데이터마이닝 기능을 추가하여 고급 분석 기능을 강화하고 있다.

그림 3-2-2 i-STREAM 워크플로 기반의 머신러닝 모델링



※ 출처: 비아이매트릭스

3. 기술 동향

빅데이터 분석을 위해 하둡 진영에서는 기존 맵리듀스 기법을 개선하기 위해 스파크가 등장하여 하둡 환경을 빠르게 재편하고 있다. 하둡 HDFS에 저장된 데이터에 SQL 질의를 쓸 수 있도록 한 SQL-on-Hadoop 제품인 아파치 타조(Apache Tajo), 클라우데라 임팔라(Cloudera Impala) 등이 등장하여 빅데이터 분석 시장으로의 접근을 모색하고 있다. 한편 몽고DB로 대표되는 NoSQL 기술은 복수의 저가 서버들을 클러스터링, 샤딩(Sharding) 등의 방법을 통해 데이터를 분리·처리하며 대량의 데이터를 빠르게 처리할 수 있는 기술로 평가받고 있다. 또한 필요에 따라 키밸류(Key-Value), 도큐먼트(Documents), 그래프 데이터베이스(Graph Database) 등으로 활용할 수 있다는 점이 특징이다.

SAP HANA로 대표되는 인메모리 DBMS(IMDBMS) 기술은 디스크가 아닌 주 메모리에 모든 데이터를 보유하고 있는 데이터베이스로 디스크 검색보다 자료 접근이 빠른 것이 장점이다. 데이터 양의 빠른 증가로 데이터베이스 응답 속도가 떨어지는 문제를 해결할 수 있는 대안으로 평가받고 있으며, 기존의 DW 환경과 기반 솔루션들을 그대로 사용할 수 있다는 점에서 강점이 있다.

클라우드 기반 분석 분야는 글로벌 IT 기업들이 딥러닝 기술을 확보하기 위해 다양한 노력을 진행 중이다. 특히 구글은 머신러닝 기술인 텐서플로(TensorFlow)를 오픈소스로 공개하여 인공지능 기술에 대한 영향력을 강화하고 있다. 고급분석 기술은 머신러닝 기법을 기반으로 해서 다양한 BI 솔루션을 접목하는 방향으로 발전하고 있다.

4. 시장 동향

국내 빅데이터 기술 및 서비스 시장 규모가 2018년까지 연평균 26.4% 성장할 것이라 전망이다. 한국IDC에 따르면 국내 주요 기업은 비즈니스 의사결정 시 적시에 정보를 얻거나 시장 상황에 신속히 대응하기 위해 빅데이터 기술을 활용하는 방안을 모색 중이다. 비즈니스 성장 기회를 식별할 수 있는 데이터 아키텍처에 대한 투자를 강화하고 있어 빅데이터 기술 및 서비스 매출이 늘고 있다는 것이다. 또한 하둡과 같은 오픈소스 빅데이터 처리 프레임워크가 비용과 확장성 측면에서 시장의 신뢰를 얻으면서 수요가 구체화되고 있다는 분석이다.

2020년까지 클라우드 기반의 빅데이터 및 분석 솔루션에 대한 투자는 엔터프라이즈 솔루션에 비해 4.5배 빠른 속도로 진행되고, 비주얼 데이터 분석 도구들은 비즈니스 인텔리전스의 다른 시장들보다 2.5배 빠른 속도로 성장할 것으로 전망된다.



5. 향후 전망

빅데이터를 위한 클라우드 플랫폼 도입과 전문 분석 서비스에 대한 관심이 높아지면서 BDaaS와 같은 새로운 빅데이터 공급체계가 부각되고 있으며, 대표적인 예가 BigML(<https://bigml.com>)이다. 빅데이터가 여전히 중요한 요소임에 분명하지만, 데이터 자체보다는 데이터를 통해 가치를 창출해내는 분석기술과 서비스에 더 집중될 것으로 전망된다. 고도화되고 예측 가능한 분석기능들이 머신러닝과 통합되어 데이터 분석 애플리케이션에 대한 새로운 성장세를 가져올 것으로 전망된다.

개발자, 데이터 과학자, 데이터 분석가들에게 직접 데이터를 탐사할 수 있는 권한을 부여해 주는 빅데이터 ‘셀프 서비스’ 환경으로 진화할 것이다. 또한 기존 분석 기능에 다양한 머신러닝 기반 기술의 접목으로 좀 더 정교하고 빠른 예측분석이 가능해질 것이다.

“데이터로 찾아야 할 행복”



2015 우수 DB인 상 데이터서비스 이노베이터 수상 / 노경운(주) **이태훈** 대표
Expert Column

바야흐로 대한민국은 데이터의 홍수다.

오늘 이 시간에도 조그마한 스마트폰에서는 날씨 정보, 취업 정보, 부동산 정보 등 다양한 데이터들이 쉴 새 없이 쏟아지고 있다. 그간 정보는 소수계층의 전유물이었으나 어느새 첨단화된 IT기기와 문화 속에서 고작 몇 번의 손가락 놀림만으로도 다양한 전문 분야의 정보를 손에 넣을 수 있게 됐다. 급기야 우리는 이런 막대한 데이터 정보로도 부족해 지금까지 다뤄 보지도, 상상하지도 못했던 천문학적 규모의 빅데이터까지 목격하고 있다.

세상은 참 편리해졌다. 어디 스마트폰 뿐이겠는가. 첨단 기술로 무장된 아파트는 또 어떤가. 우리는 아파트 거실에서 TV를 보며 쉬다가, 도시가스로 손쉽게 음식을 만들어 맛볼 수 있다. 자동제어 시스템을 통해 사시사철 온수가 나오고 추울 땐 난방이 작동하고 삼복 더위엔 에어컨이 돌아간다. 외출하기가 불편할 땐 편리한 택배 시스템으로 다양한 물건을 배달시킨다. ‘행복아파트 1004동 1004호’는 말 그대로 ‘21세기 천국의 축소판’이 아닌가 싶을 정도다.

하지만, 딱! 거기까지다.

층간소음 시비에 누가 살고 있는지 알 수조차 없는 이웃, 경비 근무자에게 막말을 해대는 ‘갑질문화’에 이르기까지, 오히려 우리는 그 ‘안락한 아파트’ 새장 속에 갇힌 주인공 신세가 되어 버린 셈이 됐다.

스마트폰은 또 어떤가. 스마트폰 게임을 하며 쉬다가 어떤 음식이 먹고 싶으면 레시피를 검색하면 된다. 아플 땐 병원 정보를 찾는다. 아이가 울 때도, 연인과 영화를 볼 때도, 집안에 필요한 물건을 고를 때도 우리들은 어김없이 스마트폰이 쥐어 주는 정보에 길들여져 살아가고 있다. 필자는 문득 아파트와 스마트폰이 21세기의 ‘괴물’이라는 생각이 들었다. 벗어나고 싶어도, 이미 셋팅된 ‘안락한’ 환경을 내가 무슨 수로 바꿀 수 있다는 말인가.

옛 고전 장자(莊子)에는 ‘물물이불물어물(物物而不物於物)’이라는 말씀이 있다. 굳이 직역을 하자면 ‘물을 물로 하고 물에 의해 물이 되지 않는다’는 조금은 아리송한 구절이지만, 좀 더 자세히 들여다보면 ‘사물을 사물로 부리되, 사물로써 사물의 부림을 받지 않는다’는 뜻으로 해석된다. 장자는 이미 수



매순간 홍수처럼 쏟아지는 데이터 속에서 생각을 가다듬을 틈조차 없다면, 넘치는 데이터가 오히려 우리의 판단과 인간다움을 해치게 될 것이다.

천 년 전에 다가올 현실을 예견했는지 모른다.

요즘 우리는 안락한 아파트와 편리한 스마트폰으로 인해 주변을 잃어버리고 결국 자기 자신마저 잊어버린 채 살아가고 있는 형국이다. 결국엔 우리가 원하는 건 안락함과 넘치는 정보는 아니었다. 아파트 생활이 아무리 편리해도 소통이 없는 주거문화가 우리에게 진정한 행복을 주기 어렵듯이 데이터 생활도 마찬가지다. 매순간 홍수처럼 쏟아지는 데이터 속에서 생각을 가다듬을 틈조차 없다면, 넘치는 데이터가 오히려 우리의 판단과 인간다움을 해치게 될 것이다.

이젠 웬만한 정보는 그냥 흘려보내자. 내가 진정으로 부족한 부분을 채울 수 있는 ‘데이터의 진정성’은 데이터의 풍부함 속에도, 또는 요즘 유행하는 빅데이터의 방대함이나 신속함 속에도 들어 있지 않은 것 같다. 다시 장자의 표현을 빌리자면 ‘데이터가 데이터를 부리되, 데이터로 데이터의 부림을 받지 않는’ 태도가 데이터를 더 진정성 있게 만들어 주는 게 아닐까 한다.

우리는 요즘 슈퍼컴퓨터가 알려주는 날씨 정보를 듣는다. 몇 시에 어느 정도의 강수량으로 얼마 동안 어느 지역에서 소나기가 내리는지 우리는 쉽게 알 수 있다. 하지만, 동시에 자연의 변덕으로 날씨 정보가 보기 좋게 빗나가 버리면 어느새 슈퍼컴퓨터를 힐난하고 날씨 정보 업체를 비난하고, 끝내는 자연까지 탓하는 우리 자신을 보게 된다. 데이터의 정교함이 인간을 더 피폐하게 만드는 꼴이다. 차라리 옛날의 ‘무딘’ 날씨 정보가 더 그리워지는 때다.

“오늘 날씨는 후덥지근하고 때에 따라 곳곳에 비가 내리겠습니다.”



제3장 데이터 컨설팅 및 구축 동향

제1절 데이터 설계 컨설팅

1. 데이터 설계 컨설팅 개요

데이터 설계는 데이터베이스 등 데이터를 관리하기 위한 체계 내에서 어떤 구조로 저장되어야 하는지 결정하는 기술적 작업이다. 반드시 데이터베이스에 저장되어야 할 데이터들만을 대상으로 하지는 않으나 어떤 형태로든 구조적 정형화를 지향한다.

국제표준화기구의 데이터 교환 및 관리표준화를 담당하는 ISO/IEC JTC1/SC32(Data Management and Interchange)에서 데이터의 의미, 구문, 표현을 표준화할 수 있는 프레임워크를 제시하기 위하여 'ISO/IEC 11179 - 메타데이터 레지스트리'를 제정하였다. 조직 내에서 관리되는 데이터의 최소 단위를 데이터 요소라고 하였으며, 정형적 구조화가 가능한 데이터들에 대해서는 데이터 요소로 표현할 수 있음을 언급하고 있다. 데이터 요소는 다음의 3개 요소로 구성된다.⁵⁶⁾

○ 객체 클래스(Object Class)

실세계의 생각, 추상 또는 사물들의 집합으로 명백한 범위와 의미로 식별이 가능하고, 데이터를 수집·저장하고자 하는 사물 혹은 개념이다.

(예) 자동차, 사람, 가정 등

○ 속성(Property)

한 객체 클래스 내의 구성요소가 가지는 일반적인 특성으로 객체를 식별하거나 분류 또는 기술하기 위해 사용되고, 객체 클래스의 내재적인 특성의 집합이다.

(예) 색깔, 성별, 나이, 주소, 수입 등

⁵⁶⁾ 국제표준화기구, ISO/IEC 11179-1:2004 Framework, 2004

○ 표현(Representation)

데이터를 어떻게 표현하는가에 대한 문제로 값 영역(Value Domain), 데이터 타입(Data Type), 측정단위(Unit of Measure), 표현 클래스(Representation Class) 등이 이에 해당된다.

즉, 정형적 구조 특성이 있는 데이터들은 객체 클래스, 속성, 표현으로 설계가 가능함을 의미한다. 데이터 모델링 시 자주 사용되는 엔티티 관계도(ERD, Entity Relationship Diagram)의 Entity, 속성(Attribute 혹은 릴레이션십 Relationship), Attribute의 도메인, 데이터타입 등이 각각 대응되는 개념이다.

표 3-3-1 데이터 구성 요소

ISO/IEC 11179 데이터 요소	ERD 구성 요소
Object Class	Entity
Property	Attribute, Relationship
Representation	Data Type, Domain 등

가. 데이터 설계 컨설팅 주요 내용

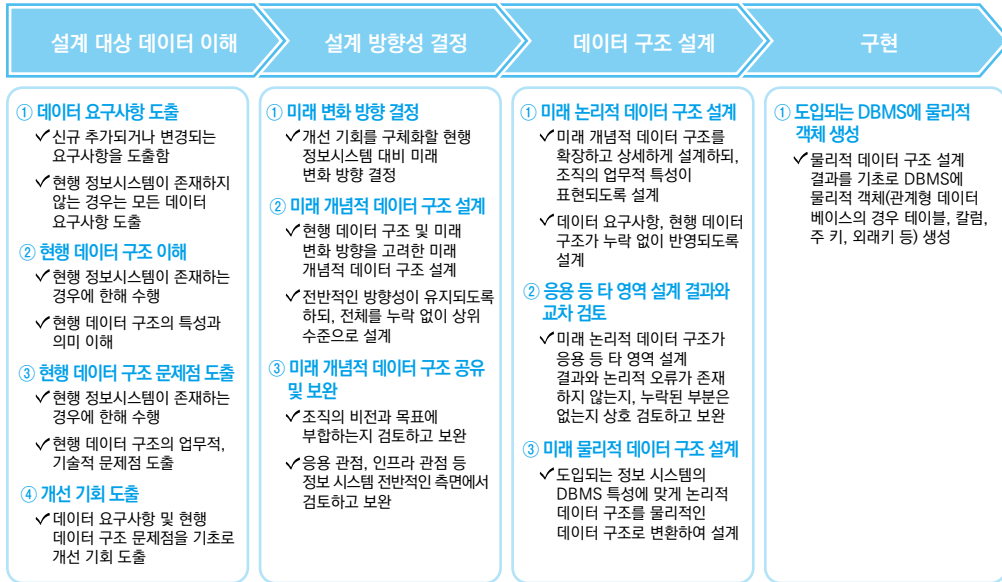
데이터 설계는 시스템의 규모에 따라 달라질 수 있기는 하지만 매우 방대한 종류의 데이터를 설계해야 하기 때문에 데이터 설계 결과에 따라 시스템 구축에서 운영에 이르기까지 응용 프로그램의 난이도, 성능, 품질 등에 미치는 영향이 크다. 나아가 해당 시스템을 이용하는 조직의 업무 효율성 및 경쟁력에도 직간접적으로 영향을 미치게 된다.

이를 다루는 응용 프로그램 설계자 혹은 개발자가 해당 영역에 필요한 요건에만 국한하여 설계하는 것은 통합적, 전체적 시각에서 많은 논리적 오류를 범할 수 있다. 따라서 전문적인 설계자에 의해 조직의 비전과 목표에 부합하는 데이터 설계를 위한 컨설팅적인 접근이 필요하다는 인식이 필요하다.

나. 데이터 설계 컨설팅 방법 및 절차

설계해야 할 데이터가 무엇인지 확인하는 과정에서부터 시작되는 데이터 모델 설계 컨설팅은 정보시스템 내에서 데이터들이 어떤 구조로 관리될 것인지를 설계하고 이를 구현함으로써 완료된다. 데이터 설계 컨설팅은 대체로 [그림 3-3-1]과 같은 흐름으로 진행된다.

그림 3-3-1 데이터 설계 컨설팅 절차



정보 시스템 사업의 특성에 따라 이러한 절차들은 일부 제외되거나 절차가 더 세분화될 수 있다. 예를 들어 정보화 전략 계획 수립(ISP, Information Strategy Planning) 사업의 경우 설계 대상 데이터 이해, 설계 방향성 결정의 전부 또는 일부만 수행할 수 있고, 정보 시스템 구축 사업 중 데이터 구조를 소극적으로 개선하고자 하는 경우 현행 데이터 구조 이해를 조금 더 강조하여 현행 데이터 구조의 문제점 중심으로만 미래 개선 방향을 수립할 수도 있을 것이다.

2. 데이터 설계 컨설팅 동향

20여 년의 시간이 흘렀음에도 여전히 관계형 데이터베이스 기반의 정보 시스템 구축 사업이 주를 이루고 있고, 이에 따라 데이터 설계에 있어서도 대부분 ERD 형식을 이용하고 있다.

전통적인 정보 시스템 구축에서 널리 진행되었던 정형화된 데이터 구조 설계 영역은 이미 대부분의 기관이나 기업이 정보화 사업을 완료하여 과거와 같이 대규모의 통합적 데이터 구조를 설계할 필요성은 많이 약해진 상태이다. 민간 기관의 경우 동일 업종의 여러 회사가 경쟁하기 때문에 특정 회사의 데이터 구조가 타 회사에 참조되고 개선되는 과정이 반복되면서 보험 등 금융산업 등에서는 상당 수준의 데이터 모델이 설계되어 운용되고 있다. 이에 반해, 행정 및 공공기관의 경우 동일 행정업무 및 공공업무를 수행하는 기관이 많이 존재하지 않고 기관마다 독립된 기능을

수행하므로 특정 기관의 데이터 모델이 타 기관에 참조되거나 개선·적용되는 경우가 드물다. 또한 더 나은 데이터 모델 확보를 필요로 하는 경쟁의 터전이 아니기 때문에 데이터 모델 품질 관점에서 민간 기관의 그것에 비해 다소 낮은 편이고 데이터 구조가 응용 설계에 종속적인 경우가 많아 정보 시스템의 경직성이 다소 높은 편이다. 이런 시각에서 볼 때 향후 민간 기업보다는 행정 및 공공기관에서의 데이터 설계 건설팅이 더 가치가 있을 것으로 전망된다. 다만 행정 및 공공기관 정보 시스템의 본질적인 개선 방향을 데이터로 이해하고 이를 위해 투자할 필요성을 어떻게 확보 하느냐가 관건이 될 것이다.

현 정부가 출범한 후 행정기관 및 공공기관이 공공데이터 개방이라는 시대적 요구에 부응해야 하는 당면과제를 받아들이고 이를 위해 공공데이터의 데이터 품질 수준 확보가 필요해졌다. 그러나 데이터 품질의 기초적인 기반이 데이터 모델이란 자각이 보편화되어 있지 않은 상태이기 때문에 데이터 품질 사업이 대체로 데이터 값 품질에만 국한되어 있으며, 값 품질 중에서도 주로 유효 값 품질 측정 수준을 크게 벗어나지 못하고 있다. 데이터 품질 측면에서 데이터 모델의 특성대로 데이터가 발생되고 있는지를 점검하는 것도 중요하기 때문에 데이터 모델 관리 여부에 따라 데이터 품질 측정 결과 지표의 신뢰성에 차이가 발생할 수 있다.

공공데이터 개방 측면에서는 기관들이 관리하고 있는 데이터가 무엇인지 알 수 있어야 개방 대상을 올바르게 선정할 수 있고, 타 기관의 개방 데이터들과 어떻게 융합되어 새로운 가치를 창출할 수 있을지 가늠해 볼 수 있다. 따라서 어떤 데이터가 어떤 구조적 특성을 지니는지를 정확하게 표현하고 있는 데이터 모델 확보는 매우 중요하다고 할 수 있다.

3. 기대 효과

노하우(Know-how)에서 노웨어(Know-where), 이제는 노후(Know-who)라는 용어도 회자되고 있다. 정보 시스템에서 정보를 어떻게 다룰 것이냐, 정보가 어디에 있는가, 누가 관리하는가라는 관점은 매우 중요하다. 특히 요즘과 같은 빅데이터 시대에는 더할 나위 없을 것이다.

그런데 여기서 간과해서는 안될 것이 바로 ‘WHAT(What)’이다. 노하우, 노웨어, 노후 이전에 노WHAT(Know-what)이 전제되어야 한다. ‘무엇’이 명확하지 않은 상태에서 방법과 위치, 주체를 정립하고자 하는 것은 주객이 전도되었다는 것이다. 정보 시스템 영역에서 ‘무엇’에 해당하는 것이 바로 데이터이다. 조직에서 다룰 수 있는 데이터가 무엇이 있고 어떤 특성이 있으며 어떻게 표현되고 있는지 정확하게 알아야 하고 이를 공통의 언어로 설계해서 지속적으로 관리하기 위한 노력이 선행되어야 이를 활용한 새로운 가치가 만들어지는 것이다.



데이터 설계의 결과물인 데이터 모델은 정보 시스템을 구축·운영하는 것뿐만 아니라 데이터를 개방하여 널리 공유하기 위해 매우 가치가 있는 국가적 자산이라는 시각이 조속히 확산, 일반화되어 좋은 데이터 모델을 설계하고 지속적으로 관리하는 것에 관심과 투자가 이어져야 할 것이다.

4. 향후 전망

여전히 전통적인 방식의 정보 시스템 구축 사업에서의 데이터 설계 컨설팅이 주를 이루고, 조직에서 쌓아둔 데이터들을 정보로 가공하기 위한 지속적인 노력도 병행될 것으로 전망된다.

한편 그간 조직 내에서 제대로 관리되고 있지 않던 빅데이터로부터 새로운 가치를 찾는 노력이 계속되고, 빅데이터와 기존 정형화되어 관리되던 레거시 데이터와의 융합 연구 및 구현을 위한 설계 요구도 꾸준히 증가할 것으로 전망된다.

제2절 데이터 품질 컨설팅

1. 데이터 품질 컨설팅 개요

국내에서 데이터 품질에 대한 관심이 시작된 2000년 이후, 현재 기업(공공기관)의 데이터 품질 관리 수준을 성숙도 단계로 볼 때 기본적인 인식단계에서 정량화(성숙도 2레벨)를 지나고 있는 수준이다. 1단계(초기) 품질활동이나 관리시스템을 도입한 기업(기관)들은 2차 단계의 고도화를 진행 중이다.

세부적으로 살펴보면 데이터표준(메타데이터)관리는 모델 구조 관리 중요성의 정착으로 필수 품질관리 요소로 자리 매김하고 있는데 비해 데이터 값 오류 추출 및 원인 분석을 통한 체계적인 통제 관리 수준은 미흡한 상태이다. 다만 데이터 측면의 개인정보 보호 법제화와 데이터 관리 소홀에 따른 민원수가 늘어남에 따라 관심 속도가 빠르고 그 목표 수준이 높아지고 있는 실정이다.

뿐만 아니라 ‘빅데이터’, ‘데이터 개방’ 시대에 기업 생존을 위해서 데이터 품질에 대한 관심과 투자를 더 이상 늦출 수 없다는 공감대가 형성되고 있으며, 이에 대한 연구와 노력이 국내외에서 심도 있게 추진되고 있다. 2013년 이후 범국가적으로 추진되고 있는 공공데이터 개방 및 활용이 현재 관리데이터의 저품질로 인해 개방추진 속도가 더딘 것도 데이터품질의 중요성을 대변해 준다.

이에 최근의 품질 컨설팅 동향을 개방데이터 품질과 빅데이터 품질 확보 측면에서 살펴보겠다.

2. 데이터 품질 컨설팅 동향

가. 데이터 개방 효과를 위한 품질관리 추진

공공데이터 활용 효과를 교통정보나 기상정보, 학교 소식, 재난 정보 등을 통해 피부로 느끼며 생활하고 있다. 하지만 아직도 많은 정보들이 수면 아래서 존재 가치를 숨긴 채 그 가치가 발휘되기를 기다리고 있다. 따라서 정부는 공급자 입장의 제한된 정보 제공에서 나아가 수요자 입장에서 실수요자가 원하는 데이터 개방(융복합 데이터 포함)을 추진 중이며, 개방데이터를 활용하기 편한 오픈API 등으로 제공방식을 개선하고 있다.

또한 공공기관 간 동일 정보인 경우 동일한 형태로 제공 및 유통될 수 있도록 표준 개방데이터 셋(전국공공화장실 표준데이터, 전국도서관 표준데이터, 전국전통시장 표준데이터, 전국도시공원 표준데이터 등)을 발굴하여 공공데이터포털(www.data.go.kr)을 통해 제공하고 있다. 뿐만 아니라 행정자치부, 각 공공기관에서 공공정보 활용 및 가치창출을 위한 ‘공공데이터 시나리오 경진대회, 공공데이터 활용 앱 개발’ 등의 행사도 추진하고 있다. 물론 이러한 목표를 이루기 위해서 각 공공기관은 데이터 개방에 앞서 기관 내부 데이터에 대한 품질향상을 위한 품질활동 관리 체계를 마련하고 있다.

그림 3-3-2 공공데이터포털의 표준 개방데이터셋

표준데이터 12건을 찾았습니다.



※ 출처: 공공데이터포털(www.data.go.kr)

나. 빅데이터 품질향상을 위한 모니터링 체계 구축 사례

한국수자원공사는 2015년 실시간 수도정보의 품질 확보를 위해 관측기기에서 발생하는 수치정보를 본부, 본사로 전송하는 과정의 데이터 품질 확보를 위해 이동 과정의 프로그램 가동 상태를 모니터링하여 적시에 원인을 분석, 재가동할 수 있도록 품질모니터링 체계를 구축하였다.

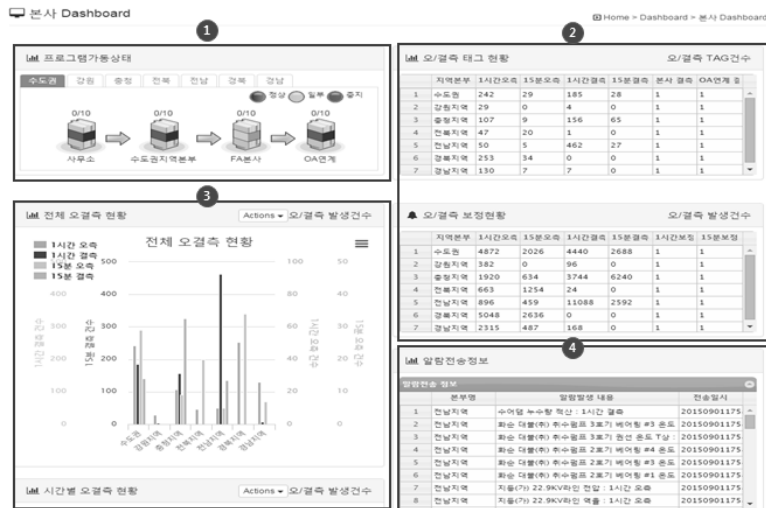
여기서 실시간 수도정보에 대한 빅데이터(관측데이터)의 특성(관측기기 상태, 환경 변화 등)에 따른 측정치의 유효값으로 인정할 수 있는 범위(MIN, MAX) 기준을 마련하여 품질관리(기준치 범위를 벗어나는 데이터는 오히려 전체 데이터 오류를 유발하여 제외)를 보다 강화하였다. 이후에도 지속적으로 증가하는 관측 태그에 대한 유효 근접 수치 범위를 확보하기 위해 머신러닝 빅데이터 기법 적용에 대한 연구를 추진하고 있다.

이전의 품질관리 컨설팅이 데이터 표준화, 정형데이터 품질 측정을 위한 업무규칙(BR: Business Rule) 도출, 전사 데이터 품질관리 체계 등의 컨설팅이었다면, 빅데이터(실시간 수도 관측 데이터)에 대한 데이터 유효범위 도출을 위한 개념과 알고리즘을 세우는 컨설팅으로 그 영역이 확대되어 가고 있는 사례이다.

[그림 3-3-3]은 실시간 수도정보에 대한 품질관리 모니터링 시스템 구축 사례이다.

- ① 데이터 처리(이동) 단계별 프로그램 가동상태 모니터링
- ②,③ 오측·결측 관측태그 수치현황 모니터링(유효범위 한계값 기준)
- ④ 오측·결측 관측태그 발생시 담당자에게 알람전송 현황

그림 3-3-3 실시간 수도정보 데이터품질관리 시스템



※ 출처: 한국수자원공사 실시간 수도정보 품질관리 체계구축_성과공유회(2015.12)

다. 데이터품질 활동체계 구축 확산

최근 크고 작은 재난 상황을 겪으면서 재난안전정보 관련 정보 통합 및 유통 관련 플랫폼 사업들이 많이 추진되고 있다. 재난은 하나의 주무기관이나 해당 지역에서만 대처할 수 있는 사안이 아니기에 가능한 국내외 재난안전 정보를 표준화된 포맷으로 적시에 유통할 수 있는 체계를 갖춰야만 미래의 재난에 대한 예측과 대비를 하고 신속한 대응과 복구를 통해 국가 자산과 국민의 안전을 보호할 수 있다. 따라서 데이터 개방과 더불어 재난안전 관련 정보의 유통표준 및 품질 등의 확보를 위한 투자와 컨설팅이 이루어질 것으로 전망된다.

한국정보화진흥원에서는 국토교통부, 농림축산식품부 등 국가 중점 DB 관리기관 중심으로 ‘공공데이터 이용활성화 지원 사업’과 ‘공공데이터 품질관리 지원 사업’ 등 개방데이터에 대한 품질활동체계 구축을 위한 품질 컨설팅지원 사업을 지속적으로 추진하고 있다. 더불어 2016년부터 본격적으로 기 지원된 기관중심으로 ‘공공데이터 데이터관리 수준 평가 기준’을 마련하여 공공기관 데이터 품질관리 수준평가를 시행하고 향후 수준평가 기관 확대를 추진하고 있다.

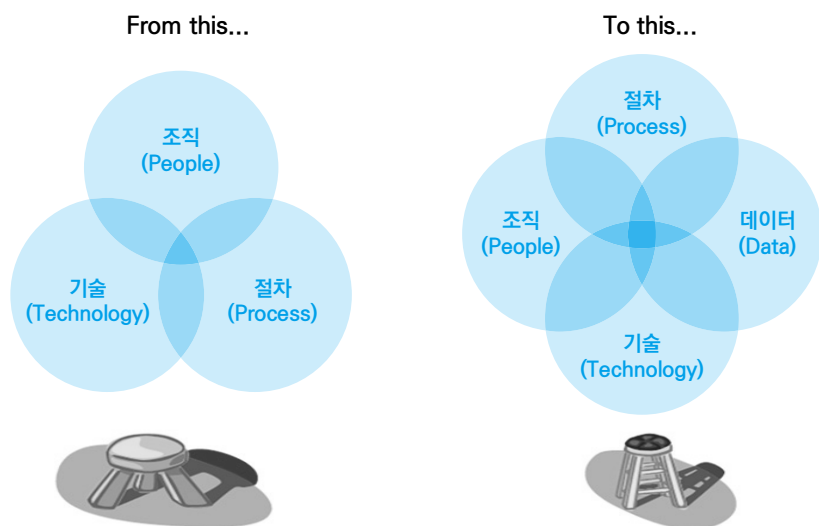
민간 영역에서도 차세대 시스템 구축시 데이터 품질관리 부문이 독립적인 한 축을 담당하여 진행되고 개발 단계에서부터 적용·운영할 수 있도록 필수 영역으로 추진하고 있다.

또한 한국데이터진흥원을 통해 데이터품질인증을 받은 기관들은 한 단계 높은 레벨의 품질인증을 추진하는 등 지속적인 데이터 품질관리 유지를 위해 노력하고 있다.

한국데이터진흥원의 ‘데이터 품질관리 교육’에 매년 많은 민간 및 공공기관 정보화담당자들이 신청하고 있어 데이터품질에 대한 관심이 날로 증대되고 있음을 알 수 있다.

2016년 미국 샌디에고에서 개최된 ‘엔터프라이즈 데이터 월드 컨퍼런스’의 한 세션인 데이터품질 발표에서 지금까지는 품질관리 핵심요소를 3개축(조직+절차+기술)으로 추진하였으나 향후에는 데이터값의 품질을 품질관리 핵심요소로 포함한 4개축(조직+절차+기술+데이터) 제시함을 볼 때 ‘데이터 값’의 품질활동 요소 강화에 대해 상기할 필요가 있다.

그림 3-3-4 데이터 품질 성숙곡선



※ 출처: 엔터프라이즈 데이터 월드 2016 컨퍼런스

3. 향후 전망

모 금융기관이 ‘ $100 - 1 = 0$ ’이라는 슬로건하에 차세대 시스템 개편을 수행하고 있는 것을 보며 지금이 ‘데이터 중심’ 시대임을 새삼 느낄 수 있다. 개방데이터 및 빅데이터의 가치를 무궁무진하게 바라보는 시각만큼 데이터 품질에 대한 시각도 그에 상응하는 품질활동과 관리 체계를 갖추는 노력이 꾸준히 이어질 것으로 전망된다.

알파고가 막대한 빅데이터를 바탕으로 대부분의 오류를 자동으로 발견하여 수정한다고 해도 ‘신의 한 수’에 당할 수 있다는 사실을 ‘알파고와 이세돌’ 간 대국을 통해 생생하게 경험했다. 데이터

/// //////////////////////////////////////

품질활동을 해야 하는 당위성이 여기에 있다. 정보시스템이 인체의 몸이라면 데이터는 혈관을 타고 다니는 피로서 온몸에 흘러다니며 GIGO(Garbage In Garbage Out) 할 것으로 저품질 데이터는 기업의 성패에 치명적인 문제를 야기할 수 있음을 잊지 말아야 할 것이다.

빅데이터 기반의 알파고 시스템이 도입되더라도 이를 맹신해서는 안 되며 이에 대한 위험성을 사전에 예측할 수 있는 품질 모니터링 알고리즘 개발에 노력해야 한다. 데이터 품질은 인간의 도덕성과 관계 관리를 통해 최상위 품질을 유지할 수 있으며 최상의 데이터를 통해 상상 이상의 가치를 발휘하게 될 것이다.

제3절 데이터 거버넌스 컨설팅

최근 들어 많은 기업들이 빅데이터 기술 도입과 분석 등을 통해 고객 경험 향상, 운영업무 개선, 새로운 가치 창출 등의 가시적인 성과를 거두고 있다.

이코노미스트의 ‘기업 경영에 있어 데이터 영향도 조사’에 따르면 데이터 활용을 통한 성과와 중요성에 대해 550명의 응답자 중 14%가 ‘데이터가 근본적인 사업 방식 변경 유도’를 선택하였고, 44%가 ‘전략적 의사결정에 데이터가 중요한 요소를 작용’한다고 답변했다. 그에 반해 데이터 품질 관리(42%), 방대한 데이터 수집 관리(34%), 데이터 보안과 프라이버시 유지(29%) 등은 데이터 활용에 장애 요인으로 나타났다.⁵⁷⁾

기업 전반에 걸친 데이터 관리 요소에 대한 점검과 보완이 요구된다는 의미로 이해 할 수 있다.

1. 데이터 거버넌스 컨설팅 개요

가. 데이터 거버넌스의 정의

기업들이 관리하는 데이터는 최근 몇 년 사이 많은 변화를 보이고 있다. 상상을 초월하는 데이터의 양적인 증가와 이전에 사용할 수 없었던 반정형 및 비정형 데이터의 활용 가능성, 기업 내부 뿐만 아니라 기업 외부 데이터까지 접근 가능한 환경, 기업이 보유한 데이터의 관리 부실로 인한 기업의 이미지 혹은 경제적 타격을 얻는 환경 등은 기업이 데이터를 통해서 더욱 가치를 발견할 수 있는 기회인 동시에 위협이 되고 있다. 이러한 상황에서 데이터 거버넌스의 접근법은 도움이 될 것이다.

데이터 거버넌스에 대한 정의와 이해는 약간씩 차이가 있으나, 데이터를 자산으로 보고 데이터 계획, 감독, 집행이라는 통제를 위한 조직·정책·절차·지침·표준 등을 수립하는 것으로 정의할 수 있다.

나. 데이터 거버넌스 컨설팅의 영역

DAMA(Data Management Association)⁵⁸⁾에서는 데이터 관리 영역을 데이터 아키텍처, 데이터 구축, 데이터 운영, 데이터 보안, 참조 & 마스터데이터, DW & BI, 문서 및 콘텐츠 데이터, 메타데이터, 데이터 품질관리, 마스터데이터 관리 등 10개로 제시하였다.

⁵⁷⁾ The Economist Intelligence Unit Limited 2015 [2015.02]

⁵⁸⁾ Data Management Association(DAMA): 1988년 설립된 데이터 관리 협회로, 데이터 품질을 비롯한 데이터 관리를 위해 노력하는 비영리 단체 기구



그림 3-3-5 데이터 거버넌스 영역



※ 출처: DMBOK2 Functional Framework, DAMA International, 2014

다. 데이터 거버넌스 컨설팅의 목적 및 효과

기업 간 경쟁이 더 이상 동종 업종에서만 발생하지 않고 업종을 넘어 전방위적으로 일어나고 있다. 빅데이터와 공공데이터의 개방으로 데이터가 넘쳐나고, 일관된 데이터 정책과 조직, 절차 등이 없인 기업 경영이 어려운 상황에 직면하고 있다.

데이터의 활용과 규제를 위배하지 않으며 경영을 해야 하는 상황에서 이제 데이터 거버넌스는 IT 부서만의 문제가 아닌 전사의 문제가 되었다.

기업 내외부 비용 절감을 비롯해 규제를 해소하고 빅데이터 등 데이터를 활용한 분석으로 시장 반응에 민첩하게 대응하기 위해서는 데이터를 자산으로 바라보는 관점과 체계적인 관리 체계 수립이 반드시 필요하다.

2. 데이터 거버넌스 컨설팅 동향 및 전망

가. 시장 규모

데이터컨설팅 시장은 2015년 1,221억원으로 2014년 대비 15.5% 성장하였으며, 각 부문 중에서 가장 큰 비중을 차지하는 영역은 데이터 거버넌스 체계 컨설팅(45.3%)으로 2015년 553억원의 시장을 형성한 것으로 나타났다.

2013년 이후로 최근 3년의 규모 추이는 데이터 거버넌스 컨설팅 영역이 데이터 컨설팅 전체 시장과 함께 지속적인 성장을 거듭하고 있다.⁵⁹⁾

표 3-3-2 데이터 컨설팅 부문별 시장 규모

(단위 : 억원)

구 분	2013년		2014년		2015년(예)		증감률	
	규모	비중	규모	비중	규모	비중	2013~2014년	2014~2015년
데이터 설계	36	3.7%	46	4.4%	45	3.7%	27.8%	-2.2%
데이터 품질	124	12.9%	135	12.8%	166	13.6%	8.9%	23.0%
데이터 관리 성능개선	34	3.5%	45	4.3%	60	4.9%	32.4%	33.3%
데이터 거버넌스 체계	488	50.6%	519	49.1%	553	45.3%	6.4%	6.6%
데이터 활용	282	29.3%	312	29.5%	397	32.5%	10.6%	27.2%
데이터 컨설팅 전체	964	100%	1,057	100%	1,221	100%	9.6%	15.5%

※ 출처: 한국데이터진흥원, 2015년도 데이터산업 현황 조사 결과 보고서

나. 데이터 거버넌스 컨설팅 전망

고객들의 서비스 기대 수준은 더 구체적인 것을 요구하고, 컨설팅 결과 또한 좀 더 가시적으로 효과를 빨리 확인할 수 있는 쪽으로 투자의 방향이 움직이고 있다. 국내 경기 침체에 따른 기업의 실용적인 투자패턴이 IT투자에도 영향을 미치기 때문이다.

데이터 거버넌스 컨설팅 시장은 데이터 거버넌스의 역할과 중요성에도 불구하고 독립적인 데이터 거버넌스 컨설팅보다는 다른 데이터 컨설팅에 접목되어 수행되는 방식이 예상된다.

따라서 향후 데이터 거버넌스 시장의 방향은 ‘빅데이터’, ‘데이터 품질’, ‘데이터 통합’, ‘클라우드’, ‘사이버보안’과 같은 최근의 IT 트렌드를 주도하는 컨설팅과 접목하거나, 그 세부 과제로 수행될 것으로 전망된다.

■ 빅데이터 거버넌스

민간, 공공부문에 빅데이터 분석 접목 및 성공 사례가 나타나고, 빅데이터 및 데이터 분석기술이 진화함에 따라 데이터 분석 시장이 확대되고 있다. 또한 기존 정형 데이터를 중심으로 한 IT 개발 및 운영 관점의 데이터 거버넌스 체계를 보완하는 요건이 발생하고 있다.

예를 들어 빅데이터 분석 지원을 위한 데이터 아키텍처 마련, 내·외부 비정형 메타데이터 관리, 분석업무 지원을 위한 데이터 목록 검색 등 빅데이터 분석시 데이터 거버넌스 관점에서 고려해야 될 요소들이 예상된다.

59) 한국데이터진흥원, 2015년 데이터산업 현황 조사 결과 보고서, 2015.12

표 3-3-3 빅데이터 거버넌스 주요 고려 요소

부문	고려 요소
데이터 아키텍처	• 빅데이터를 포함한 분석 요건 지원을 위한 기존 정보 분석 환경에 빅데이터 요소 기술 접목 및 활용 방안 수립
메타데이터	• 정형 정보 중심의 IT 운영 지원 관점의 메타데이터 관리 환경에서 빅데이터 환경의 메타 정보, 흐름 정보 등 수집 및 제공 체계 구축 • 데이터 과학자 및 분석가용 데이터 검색 체계 구축
마스터데이터	• 기업 내·외부에 산재된 각종 정보(예: 웹사이트 로그, 소셜 활동 내역 등)를 고객 마스터 중심으로 연계 및 활용 체계 구축
데이터 생명주기 관리	• 급격히 증가하는 내·외부의 정형·비정형 데이터 활용을 고려한 데이터 백업, 폐기 방안 수립
데이터 품질관리	• 빅데이터 기술 환경에 대한 데이터 품질 측정 및 클렌징 방안 • 분석 요건에 적합한 수준의 데이터 품질 관리 방안
보안 및 프라이버시	• 데이터 공유 및 활용 확대에 따른 개인·민감 정보 식별 및 관리 • 개인정보 사용 인가 관리 및 조회 모니터링 방안

빅데이터 거버넌스 컨설팅 수요의 증가는 구체적인 기업의 요구에 기반하지 못할 경우 투입된 비용 대비 효과적인 결과를 기대하기 어려울 수 있다. 따라서 빅데이터 거버넌스 컨설팅을 수행하는 수행사와 컨설팅 서비스 받기를 원하는 고객은 [표 3-3-3]과 같은 요소를 사전에 고려할 필요가 있다.

■ 공공데이터품질 거버넌스

지속적인 공공데이터 개방 정책 추진과 이에 따른 민간부문 활용 확대가 이루어지고 있으나, 실제 민간부문에서 데이터 활용시 개방 데이터에 대한 정확성, 적시성, 지속성 관점에서 다양한 데이터 품질 문제가 발생하고 있다. 개방 중인 공공데이터 품질 문제는 공공데이터 활용 확대에 핵심적인 저해 요인으로 대두되고 있다.

이를 위해 행정자치부는 공공데이터 품질관리 수준 평가 제도를 준비하여 2016년 5월 민간 활용도가 높은 21개 분야의 공공데이터를 대상으로 '품질관리 수준 평가'를 실시할 것이라고 밝혔다.⁶⁰⁾ 이에 따라 그동안 공공데이터 이용 활성화 지원사업과 공공기관의 자체사업을 통한 규범적 데이터 품질관리 체계수립 컨설팅 시장은 데이터의 실질적인 활용도와 해당 개방기관의 업무적·데이터적 특성을 고려하여 데이터 문제점을 실질적으로 개선 및 유지하는 쪽으로 초점을 맞추어 공공데이터 품질 컨설팅 시장이 변화될 것으로 전망된다.

60) 2016.2.1.(월) '공공데이터 전략위원회'(공동위원장: 황교안 국무총리, 김진형교수) 제6회 회의 심의·확정 발표, 「2016년 국가중점데이터 개방 추진계획」, 「공공데이터 품질관리 추진계획」, 「공공 빅데이터 운영 활성화 추진계획」

■ 고객데이터통합 컨설팅

금융 부문의 고객 및 상품, 제조·유통 부문의 제품 및 부품 등이 전통적 마스터데이터 관리 컨설팅 시장이었다. 여기에 빅데이터 및 고객경험 분석이 확대되면서 기업 내·외부에 산재된 고객정보에 대한 관리기준 수립과 통합 체계 마련 필요성이 증가되었다. 따라서 고객 데이터 통합(Customer Data Intergration) 컨설팅 시장이 확대될 것으로 예상된다.

고객 데이터 통합 컨설팅은 ‘수집’ → ‘정제 및 표준화’ → ‘데이터 보강(Data Enrichment)’ → ‘통합’ 과정 상에 고객 데이터 항목별 데이터 프로파일링, 정제 및 표준화 기준 정립, 데이터 보강 방안 수립, 복수 데이터 원천의 동일 데이터 항목에 대한 통합 우선순위 기준 정립 등이 세부 컨설팅 과제로 수행될 것으로 예상된다.

3. 데이터 거버넌스 컨설팅 방향성

첫째, 거버넌스 컨설팅은 특정 정책결정자 중심으로 시작은 가능하지만 조직 전체에 대한 이해와 필요성에 공감 과정이 없다면 거버넌스의 추진 목적은 추상적인 수준에 머물거나 전체 조직원들과 유리되어 좌초될 수밖에 없다. 이러한 문제들은 데이터 거버넌스 수립의 필요성을 체감할 수 있는 과제 중심으로 비즈니스 문제를 발굴하고 해결하는 방식이어야 초기의 추진력이 지속적으로 유지될 수 있다.

둘째, 비즈니스 환경 변화에 따라 주된 이슈가 되는 빅데이터, 클라우드, IoT 개인정보 보호, 공공정보 개방, 고객통합, 마스터데이터 통합 등은 구체적인 비즈니스 문제와 결합되어 데이터 거버넌스가 수립될 때 거버넌스의 수립 성과를 담보할 수 있다.

제4절 데이터 이행 컨설팅

1. 데이터 이행 컨설팅 개요

국내 IT산업은 1980년대 메인프레임 시대와 1990년대 유닉스 오픈시스템의 부흥기를 거치면서 비약적인 발전을 거듭하였다. 차세대라는 이름하에 금융, 통신, 제조, 서비스 등 업종을 불문하고 정보처리 시스템이 재구축되기 시작했으며 인하우스 개발부터 패키지 구축사업까지 SI산업의 눈부신 발전이 지금까지도 지속되고 있다.

하드웨어산업의 놀라운 발전 속도와 하드웨어 가격 하락 등에 힘입어 IT업계는 시대적 트렌드라 할 수 있는 하나의 흐름을 타고 지속적으로 변천해 왔다. 예를 들어 1990년대 초는 메인프레임 시대 이후 다운사이징이 유행하면서 유닉스 오픈시스템이 대세를 이루며 클라이언트 서버 형태의 분산시스템이 유행하였다. 이후 1990년대 후반에 접어들어 미들웨어시장이 확장하며 소위 말하는 3계층구조(3-Tier Architecture)가 보편화되기 시작하면서 인터넷과 인트라넷으로 대별되는 인터넷컴퓨팅 시장으로 진입하게 된다. 이러한 IT산업의 진보는 IT시스템이 단지 현업의 업무를 보조해 주는 측면에서 벗어나 기업의 핵심 시스템으로 자리잡게 되었다. 2000년대 접어들면서 고가 용성, 확장성, 고성능, 사용 편의성 등이 핵심적인 요구사항으로 부상하게 된 이후 현재는 빅데이터라 불리는 방대한 데이터의 분석을 통해 기업 가치를 찾는 데이터 시대로 접어들게 된 것이다.

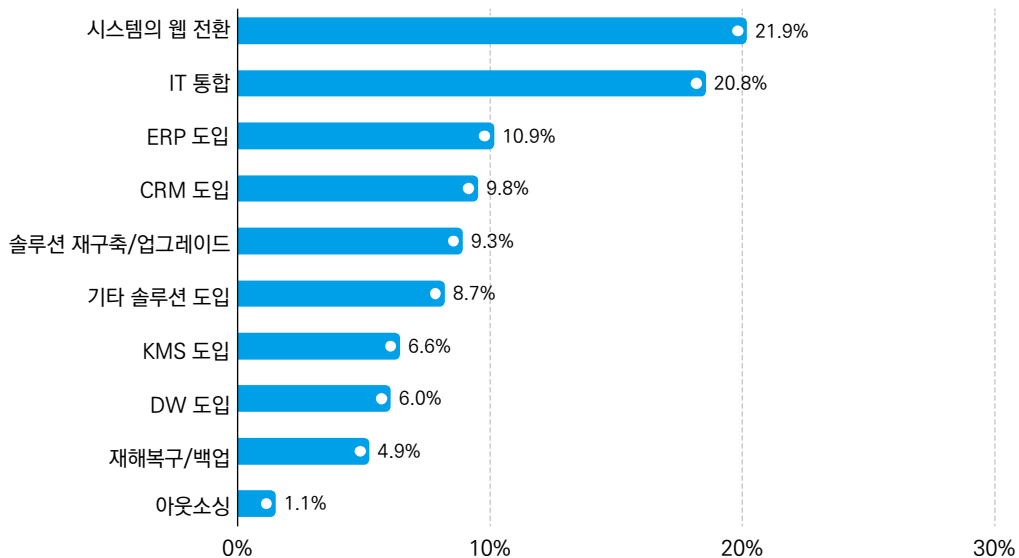
이렇게 시스템의 변천이 있을 때마다 ‘데이터의 이행’이라는 행위는 지속적으로 요구되었으며 그 과정에서 오랜 기간 동안 축적된 경험과 노하우를 바탕으로 데이터 이행 방법론이라는 하나의 컨설팅 체계를 갖추게 되었다. 즉 더 좋은 성능의 하드웨어, 스토리지의 출현과 차세대 시스템 등에 힘입어 새롭게 개발되는 시스템의 증가와 그에 따른 데이터의 통합 및 이관 등으로 기업의 핵심 자산인 데이터 및 스키마 정보들이 새롭게 통합, 이관되고 신시스템에서 효율적으로 사용될 수 있도록 하는 일련의 행위를 데이터 이행 컨설팅으로 정의할 수 있다.

2. 데이터 이행 컨설팅 동향

기업의 데이터는 전환(Migration)과 통합(Integration)이라는 이름으로 지속적이고도 반복적으로 진화를 거듭해 왔다. 국내 IT 리서치회사인 KRG에서 기업의 투자 우선순위를 조사한 바에 따르면 ‘시스템의 웹으로 전환과 IT 통합’ 부문이 42.7%를 차지할 정도로 높은 관심을 받고 있는 것으로 나타났다. 최근 10여 년간 급변하는 정보사회에서 기업 경쟁력 향상을 위한다는 명목으로 체계성 없이 시스템들이 개발되면서 IT 예산은 기하급수적으로 증가되었다. 그러나 최근에는 이러한 IT 투자에 효율화를 기함으로써 새롭게 기업 경쟁력을 강화하는 방향으로 변화하고 있다. 실제

로 그 동안 기업은 IT 신기술 도입에 많은 투자를 했지만 비체계적인 시스템 도입으로 인해 막대한 유지 보수 비용에 시달려 왔다. 특히 IT 공급자의 유지보수 비용 인상은 기업의 IT 비용 부담을 더욱 가중시키고 있다. 뿐만 아니라 데이터 및 사용자 증가로 인한 시스템 성능저하, 시스템 자원 증설의 한계와 오픈시스템 등 신기술 접목 및 기술 진화에 대한 한계 등 다양한 문제점에 봉착하게 되었다. 이와 같은 문제점들을 해결하고자 기업들은 지속적으로 시스템 통합 및 업그레이드를 고민하게 되었으며 이 과정에서 자연스럽게 데이터 마이그레이션의 필요성이 대두되었다. 데이터 이행 시장은 크게 두 가지로 구분할 수 있다. 첫째 기존 시스템을 통합하고 신기술로 시스템을 업그레이드 또는 고도화하는 것과 둘째 유지 보수 비용을 절감하고 안정성이 보장된 IT 제품으로 교체하는 것이다.

그림 3-3-6 기업 애플리케이션 투자 우선 순위



※ 출처: KRG

가. 데이터베이스 시장의 변화

기존에 사용하고 있는 하드웨어가 노후되었거나 리스가 만료된 경우 새로운 환경에 맞춰 하드웨어를 교체하는 시점에 새로운 공급자를 선정해야 하며, 낡은 소프트웨어를 새로운 환경에 적합한 소프트웨어로 교체하는 작업이 필요하다. 기업은 주기적으로 새로운 환경에 적합한 하드웨어를 선정해 교체하는 작업을 해야 하며, 프로그램 언어도 바뀌어야 한다. 하드웨어, 데이터베이스, 미들웨어 등의 변화는 매우 느린 반면 프로그램 언어는 매우 다양하고 복잡하게 변화하고 있기 때문이다.

오라클은 이미 미들웨어 업체인 BEA 사를 인수 합병해 미들웨어와 데이터베이스 시장에서 큰 반향을 일으킨 바 있으며 경쟁 업체들은 오라클의 이러한 변화에 대처하고 있다. 한국IBM은 DB2와 웹스피어를 내세워 시장에 적극 나서고 있다. 마이크로소프트, 사이베이스, 테라데이터 등도 시장에 깊숙하게 참여하고 있다. 나아가 티맥스 등 국내 기업들도 데이터베이스와 미들웨어 시장에 뛰어들어 외산 공급업체들과 극심한 경쟁을 펼치고 있다. 최근에는 정부 공공시장을 중심으로 가격 대비 성능 측면에서 국내 기업 제품들이 외산 제품들에 뒤지지 않는다는 것이 입증됨에 따라 데이터베이스 시장의 지각변동이 예상되고 있다. 이렇듯 시장에서 지각변동이 일어나면서 DBMS 시장에서도 데이터 이행에 대한 전문성의 요구와 관심도가 높아지고 있다.

나. 데이터 이행 방법론

고객들은 다양한 환경에서 데이터 이행을 수행하게 되는데 컨설팅 업체나 전문물 제조사 등에 따라서 다양한 방법론이 소개되고 있지만, 공통적으로 다음과 같은 단계로 수행하게 된다.

1) 이전 계획 수립 및 분석단계

① 용량계획(Capacity Plan)

기존 시스템의 리소스를 파악하며 시스템 간의 연동 및 트랜잭션을 분석하고 각 Product에 대한 신 기술을 분석 후 각 업무에 대한 접목대상을 선정하는 등 신규 시스템에 대한 구성방안을 설계하는 단계이다.

② 이행계획(Migration Plan)

데이터 이행 관련 마스터플랜을 설정하는 단계이다. 세부적으로는 이관 대상OS, DBMS 등 제반 소프트웨어의 호환성 검사를 통하여 개발 환경을 구성해 주며 실질적인 데이터 이행 방안 및 데이터베이스 테이블별 이관 시점, 비상시 복구방안 등을 마련하게 된다.

2) 설계/개발 및 환경구축단계

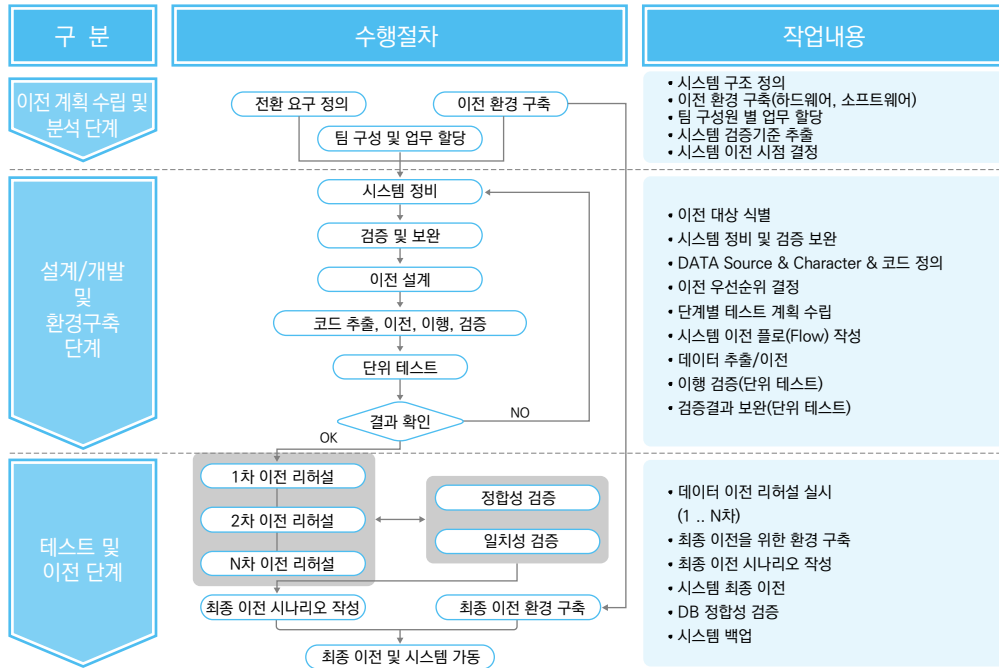
목표 시스템의 스토리지 환경 구성 점검 및 논리적 DB 환경 구성을 확인하며 시스템 모니터링 및 비상 복구 계획 등을 수립한다. 또한 이전 우선순위를 결정하고 단계별 테스트 계획을 통한 단위 테스트를 실시하고 그 결과를 검증 보완한다.

3) 테스트 및 이전 단계

데이터 이행의 리허설을 실시하며 그 결과에 따라 최종 이전을 실시한다. 이행의 진행은 테스트

및 이행 종료시까지 데이터의 정합성, 일치성 등을 반복적인 검증을 통하여 수행하는 단계로서 이행 전후의 트랜잭션 성능도 반드시 비교평가하게 된다.

그림 3-3-7 데이터 이행 절차 및 내용



3. 고려 사항 및 기대 효과

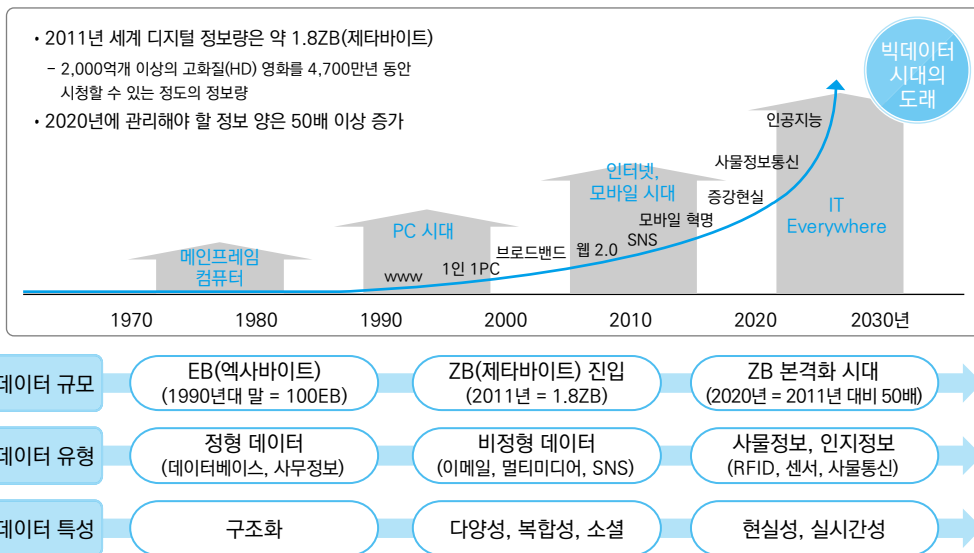
가. 데이터 이행 컨설팅의 주요 고려 요소

1) 대용량 데이터 (VLDB, Very Large DataBase)

데이터 이행을 요청하는 고객사는 원장 데이터는 물론 트랜잭션 데이터와 그에 따른 로그 데이터 등 이행 대상으로 정의된 데이터가 상상 이상으로 다양하고 방대하다. 특히 금융권이나 물류, 장치산업 등에서는 이력데이터 및 로그데이터의 경우 초당 수~수십기가 바이트 이상의 데이터가 누적되고 있어, 이러한 데이터의 무중단 이행이란 쉽지 않을 뿐더러 이행전략을 수립하는 단계에서부터 치밀한 전략과 효율적인 이행 방안이 수립되어야 한다. 한편 과거에는 소위 미션 크리티컬 분야로 분류되지 않았던 DW 데이터 부문도 현시점에서는 더 이상 그렇지 않게 되었다. 1990년대 초반부터 태동하기 시작했던 데이터웨어하우스는 20여 년을 지나오며 많은 기업의 정보계 핵심시스템으로

자리를 잡기 시작했으며 수분 수초의 장애도 RTE(Real Time Enterprise) 기업경영에 심각한 문제를 야기할 수 있는 상황이 현실이 된 것이다. 바로 이 DW데이터가 또 하나의 대용량 데이터라는 관점에서 이행정선텔의 주요대상이 되고 있다. 특히 전사 정보지원의 인프라가 되는 EDW(Enterprise Data Warehouse)의 경우 국내의 기준으로 보더라도 기본적으로 수십에서 수백 테라바이트를 넘어 이제는 페타바이트급의 초대용량 데이터규모로서 이행정락을 수립하는 과정에서 기존 OLTP 시스템과의 연관성을 고려하여 이행 순서 및 방법을 차별화하여 고려해야 할 것이다.

그림 3-3-8 세계 정보량 증가 추이



※ 출처: IDC&EMC, 디지털 유니버스 스터디, 한국정보화진흥원, 2011

2) 비정형 데이터의 증가

10여 년 전까지만 해도 데이터 이행이란 RDBMS 등 정형화된 데이터구조에 저장된 데이터를 대상으로 하는 것이 대부분이었으나 요즘에는 로그 데이터를 포함하여 콜센터 등에서 생산되는 음성 녹취파일, 이미지 영상파일, 빅데이터에서 활용되는 하둡 데이터 등 기존 RDBMS와 같은 정형화된 표준데이터 체계를 벗어난 비정형 데이터가 이행의 대상으로 포함되고 있다. 이는 DB 벤더사나 서드파티 솔루션으로 출시되고 있는 이행도구들도 미처 포함하지 못한 것으로 지금까지도 수많은 고객의 다양한 형태로 존재하는 비정형 데이터를 어떻게 이행할 것인가에 관한 요구사항으로 남아 있다.

3) 무중단 이행

과거에는 고객사가 데이터 이행계획 수립 시 최대 수 시간의 다운타임(Down Time)을 허용하였던 시절이 있었다. 수행사는 주어진 시간 내에 반드시 데이터를 이관하는 계획을 수립하여왔다. 그러나 최근 들어서는 정보시스템 고가용성의 중요성이 필수 요건이 되면서 무중단 또는 시스템 리부팅 정도의 시간인 1시간 미만 등 최소한의 시스템 중단 시간만이 허용되는 현실이다. 무중단이든, 1시간이든 주어진 다운타임이라는 것은 작업을 수행하는 이행사업 수행자의 입장이 아니라 철저히 고객사 입장에서 정해진 시간이기 때문에 만일 주어진 시간 내에 성공적으로 이행을 수행하지 못한다면 이행시작 시점의 데이터 상태로 원상복구하여야 하며 또한 그 사이에 진행된 트랜잭션까지 반영하여 데이터의 무결성을 보장해야 하는 역이행 작업을 수행해야 한다. 물론 이행 시간 내에 작업이 수행되지 못한 원인이 명확하고 원복 절차가 완벽하여 원복시 정확하게 데이터 정합성이 보장되는 경우라면 다행이지만 대부분의 경우 생각지도 못했던 원인에 시간만 소비하며 작업이 중단되고 원복 절차도 정확하지 않아 결국 데이터 무결성이 깨지거나 일부 유실되는 상황까지도 발생하는 것을 종종 볼 수 있다. 이는 데이터베이스의 특징상 데이터베이스를 둘러싸고 있는 복잡한 주변환경의 이해 및 정리가 완벽하지 못하여 발생하는 현상으로써 데이터 이행 시 고려해야 할 사항들을 간과해서는 안 된다. 무중단 데이터 이행을 위하여 고려해야 할 사항들은 다음과 같다.

① 정확한 이행시간 예측

철저한 이행계획과 반복적인 모의실험 등을 통하여 정확한 이행시간을 예측해야 하며 고객과 협의할 예상이행 시간은 이보다는 최대 50% 여유있게 설정해야 한다. 어떤 솔루션을 활용하여 이행작업을 하는가에 따라 다르겠지만 대부분의 경우 이행작업이 길어질수록 데이터의 정합성은 비례하여 어려워지며 원복하게 될 경우를 대비해야 하기 때문이다.

② 전체적인 시스템 이행과제의 이해

데이터 부문만으로 볼 때는 문제없이 성공적인 이관이 완료되었다고 하더라도 데이터베이스는 WEB, WAS를 포함한 구동 애플리케이션들의 이관과 맞물려 구동되기 때문에 데이터 관점을 넘어선 시스템 전체의 이행전략을 면밀하게 이해해야 한다. 즉 부문별 이행팀 간의 충분한 협의와 철저한 사전준비를 통해서 시스템, 애플리케이션 등과 연계하여 고려해야만 최종적인 데이터 이행의 성공을 담보할 수 있기 때문이다.

4) 원복절차 및 사각지대 트랜잭션 데이터 복구방안 확보

데이터 이행이 실패하였거나 주어진 시간 내에 완료하지 못하게 될 경우 계획된 시나리오대로 원복절차에 착수하게 된다. 이 경우 흔히 간과하는 경우가 트랜잭션 데이터의 처리 문제이다. 이를 해결하기 위해서는 데이터베이스의 특성을 정확하게 이해하고 활용 가능한 복구방안을 정확하게 준비해야 한다.

5) 검증된 이행 솔루션의 선택

시장에는 하드웨어 및 소프트웨어 솔루션을 막론하고 다양한 무중단 데이터 이행 솔루션들이 출시되어 있다. 정확한 이행전략 및 계획이 있다면 굳이 매뉴얼 작업이나 스크립트 작업에만 의존할 것이 아니라 적절한 솔루션을 활용하는 것도 매우 중요하다. 이 경우 이행전략 수립 단계에서 개념 증명(POC, Proof of Concept)과 벤치마크 테스트를 통한 사전 솔루션 검토가 필요하다.

6) 이행 결과 검증 및 정합성 확보

정형·비정형의 대용량 데이터를 무중단으로 이행 완료하였다고 했을 때 그 성공 여부는 어떻게 판단하게 될까? 이는 데이터 이행 작업 자체만큼이나 중요하고 필수적인 요소다. 이행 결과의 검증 절차는 이행 프로젝트 규모에 따라서 다소 차이는 있겠지만 고객사 담당자의 개입이 그 무엇보다 중요하다. 즉 데이터의 검증 절차는 프로젝트 초기 데이터 이행 전략 수립단계에서부터 정의되고 현업담당자와 협의된 규칙에 따라 검증프로그램이 작성되어야 하는 것이다. 데이터의 정합성을 보장하기 위한 데이터 검증 주요 이슈는 다음과 같다.

- ① 데이터 검증 절차 및 방법의 명확화
- ② 데이터 검증시 발견된 오류에 대한 원인 규명과 조치 및 이에 대한 모니터링 체계 확보
- ③ 변환 및 오류 데이터 처리에 대한 신속한 의사결정
- ④ 현업 및 관련 팀과의 유기적인 협조
- ⑤ 데이터 검증 시 최종 데이터 적재를 위한 볼륨 및 적재 시간 테스트 반복 실시

7) 이행 이후 튜닝

이제 데이터 이행 프로세스 중 마지막 절차인 이행 후 튜닝에 관하여 알아보기로 한다. 예를 들어 기존 집에서 새집으로 이사를 갔다고 가정해 볼 때 기존 가구들이 새집 구조에 따라서 어떻게 배치되는 것이 공간 활용에 좋을지, 또는 사람들의 동선이 어떻게 변하면 좋을지 등이 이에 해당한다.

DB 서버 스펙(CPU, Core수, 메모리 사이즈 등)과 병렬 구성 여부 그리고 OS 버전 및 특징, 스토리지의 구성, 네트워크 밴드위스 및 구성 등 데이터베이스를 둘러싼 하드웨어 및 네트워크, 운영소프트웨어의 변화에 따라서 데이터의 연산 성능이 기존 성능보다 개선되지 못한다면 낭패가 될 수 밖에 없다. 또한 DBMS 자체로만 볼 때도 버전이 바뀌고 구성된 물리모델도 바뀌게 된다면 성능 문제가 그 무엇보다 민감하게 반응하게 되므로 DBMS 전문가의 튜닝은 반드시 필요하다. 통상적으로 데이터 이행을 하게 되면 이행 이전의 기존시스템 성능을 100으로 본다면 120에서 130은 보장해야 한다. 즉 30% 이상의 성능 개선이 담보되어야 전반적인 이행 만족도를 보장할 수 있다.



4. 향후 전망

데이터 이행 컨설팅 및 이관/통합기술은 IT시장의 발전과 함께 성장하였으며 고객의 다양한 요구사항을 충족시키기 위하여 진화해 왔다. 그 여파로서 데이터 이행 산업은 유관 컨설팅 시장을 비롯하여 솔루션, 톨시장과 수행 DBA, 개발자들을 필요로 하는 또 하나의 파생시장으로 그 위상을 키워 왔다.

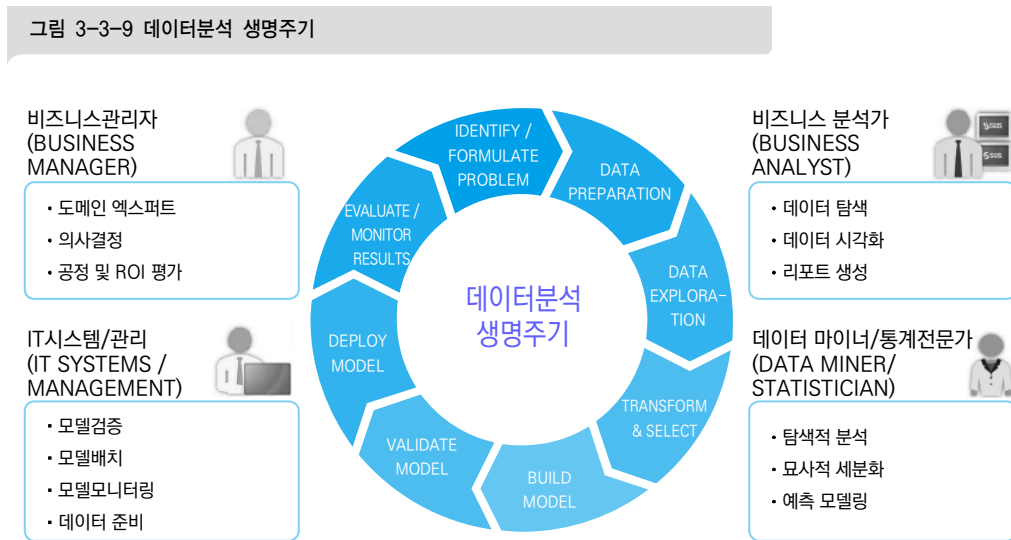
데이터 이행 컨설팅 시장은 ‘구 시스템으로부터 신 시스템으로의 효과적인 데이터의 이행’이라는 본연의 목적을 달성하는 것 자체로서도 중요한 의미가 있지만, 이 과정에서 파생되는 데이터의 정제 등 데이터 품질 향상에도 주요한 역할을 하였다. 또한 차세대 시스템의 높은 설계 품질을 요구하며 더 나아가 데이터 거버넌스 구축을 유도하는 등 2차 파급효과를 창출하고 있는 것 또한 사실이다. 이러한 고유의 목적과 장점으로 향후 시장 전망도 긍정적이다.

다만 데이터 이행 시장 증가 규모 면에서 볼 때 클라우드가 향후 IT 트렌드로 평가되고 있는 바와 같이 온프레미스 환경의 시스템들이 클라우드 시스템으로 이행되는 과정에서 단기적으로는 증가할 것으로는 예상되나 궁극적으로 PaaS(Paltform as a Service) 구축 환경이 안정화되고 기업의 온프레미스 시스템 자체가 감소함에 따라 장기적으로는 데이터 이행 시장도 단계적으로 감소하는 추이를 보일 것으로 전망된다.

제5절 데이터 분석 컨설팅

1. 데이터 분석 컨설팅 개요

데이터 분석 컨설팅 영역은 비즈니스 의사결정에서 필요로 하는 정보를 데이터로부터 생성하는 과정이라고 정의할 수 있다. 이 과정은 일반적으로 데이터를 준비하는 과정, 탐색적 데이터 분석 과정, 모델링 과정, 검증 및 피드백 과정으로 구성된다. 데이터 분석 컨설팅의 일반적인 절차는 [그림 3-3-9]와 같다.



※ 출처: SAS

가. 데이터 준비 단계

데이터 준비 단계는 정보 생성에 필요한 데이터를 통합하는 과정으로, 데이터 요건 분석, 데이터 수집, 데이터 정제 과정을 거쳐 통상적으로 분석 전용 데이터 마트를 구성하게 된다. 분석 데이터 마트에는 내부 데이터와 다양한 외부 데이터들이 핵심 분석 단위를 기준으로 생성된다.

나. 탐색적 데이터 분석 (Exploratory Data Analysis)

탐색적 데이터 분석은 데이터의 특징과 내재된 구조에 대한 인사이트를 찾는 과정이다. 이 과정의 핵심은 어떤 기술적 방법론들을 사용할 것인가의 문제가 아니라, 데이터 분석을 수행하기 위한 자재와 철학의 문제이다. (어떻게 데이터의 세부 항목을 볼 것인가? 분석의 목적이 무엇인가? 어떻게 분석할 것인가? 어떻게 결과를 해석할 것인가? 등) 탐색적 데이터 분석 과정에서는 '통계적

그래픽'이 가장 많이 사용되며, 데이터의 기본 특성 분석, 데이터의 발생 메커니즘에 대한 분석, 중요한 변수의 판별, 비정상적인 데이터의 구별 등에 대한 절차를 생성하고 필요한 경우 데이터 마트에 반영한다.

다. 모델링

모델링은 데이터가 포함하고 있는 특성을 함수 형태의 결정론적인 요소와 오차항에 의한 요소로 구분하고 데이터의 변화를 표현하는 함수를 추정하는 과정으로, 주로 추정·예측·최적화에 대한 모형을 생성한다. 특히 예측과 최적화 모형이 많이 사용되고 있다.

최근 들어 전통적인 통계 방법론과 더불어 머신러닝이 세간의 관심을 받고 있다. 머신러닝은 데이터 분석 모델 구현을 자동화하기 위한 방법론의 일종으로 반복적인 학습 알고리즘을 이용하여 데이터에서 인사이트를 정확하게 찾을 수 있게 하는 방식으로, 흔히 '기계가 스스로 학습한다'라고 표현하고 있다.

2. 데이터 분석 컨설팅 동향

데이터 분석 컨설팅 영역은 분석 기반 플랫폼의 변화와 컴퓨팅 환경의 변화에 따라 매우 빠르게 변화하고 있다. 특히 최근 몇 년간 '빅데이터 플랫폼'의 구축이 활발하게 진행되어 대부분의 공공기관들의 경우에는 빅데이터 플랫폼 구축을 어느 정도 완성한 상황이다. 또한 컴퓨팅 환경도 클라우드 컴퓨팅 환경의 구축 및 다양한 병렬처리 환경, 인메모리 기반의 분산컴퓨팅 환경 등이 대두되면서 데이터 분석 컨설팅 영역에도 많은 변화를 가져오고 있다.

가. 빅데이터 플랫폼의 구축

빅데이터 플랫폼의 구축은 데이터 분석 환경에서 좀 더 다양한 데이터를 활용할 수 있는 환경을 제공하고 있다. 기존의 내부에서 관리하던 데이터와 더불어 외부 데이터를 쉽게 활용하여 분석할 수 있는 환경이 주어지면서 데이터 분석 영역에서도 '빅데이터'에 대한 모델링 방법론, 이미지, 텍스트, 음성 데이터 다양한 형식의 데이터 분석 방법론, 스트리밍 데이터에 대한 분석 영역 등이 새롭게 부각되고 있다.

나. 컴퓨팅 환경의 변화

컴퓨팅 환경의 변화는 데이터 분석 컨설팅 영역에 더 큰 영향을 주고 있다. 빅데이터 분석과 더불어 제공되는 다양한 환경(클라우드, 인메모리, 인-DB 등)은 더 복잡하고, 다양한 데이터에

대한 더욱 정확한 분석을 가능하게 하지만, 이를 수행할 수 있는 더 많은 인력을 필요로 한다.

다. 머신러닝

구글의 ‘알파고’가 기존에는 불가능할 것이라고 평가되던 바둑에서 인간의 영역을 넘어서는 결과를 보여준 후 머신러닝에 대한 관심은 급격하게 증가하여, 실제 비즈니스에 접목하기 위한 시도가 진행되고 있다.

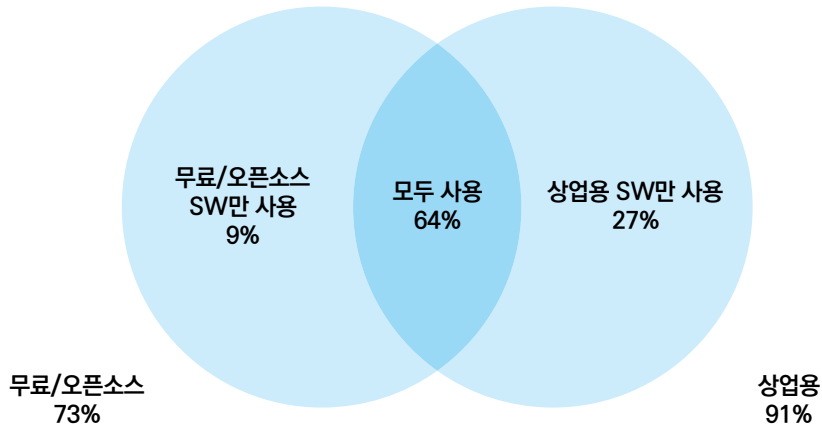
최근 머신러닝의 활용은 매우 다양한 분야로 확대되고 있다. 금융권의 ‘사기적발’ 모형의 개발과 고객관리 업무 적용을 위한 ‘실시간 마케팅’, ‘텍스트마이닝’, ‘소셜 미디어 분석’, ‘신용평가 모형’ 등과 제조업의 스마트 팩토리 사업의 일환으로 ‘불량분석’ 모형의 개발이 더욱 확대되고 있다.

라. 오픈소스 환경의 부각

기존 데이터 분석 영역이 대부분 상업용 소프트웨어를 기반으로 이루어지고 연관 시스템도 상업용 솔루션을 중심으로 되었으나, 최근 들어 오픈소스 R과 파이썬을 활용하여 데이터 분석 환경을 구축하는 사례가 급격하게 증가하고 있다. 대부분의 공공 사업은 오픈소스 활용을 주요 사업 규격으로 제시하고 있고, 특히 머신러닝 분야의 경우 오픈소스 기반의 솔루션들이 기존 상업용 솔루션과의 경쟁에서 우위를 점하는 현상이 발생하고 있다. 특히 오픈소스 R을 기반으로 하면서, 분산환경, 병렬처리환경, 인-DB 분석을 지원하는 솔루션들이 등장하면서 R에 대한 관심은 더욱 커질 것으로 보인다.

실제로 분석 및 데이터 공학 컨설팅 기업 KD너겟(KDnuggets)의 2015년도 조사 자료에 따르면[그림 3-3-10], 데이터 분석 영역에서 상업용 소프트웨어만 사용하는 비율은 27%에 그치고 있으며, 상업용과 오픈소스를 동시에 활용하고 있는 비율은 64%로, 2014년도 조사의 49%에서 약 30% 정도 증가하였다.

그림 3-3-10 오픈소스와 상업용 소스 활용도



※ 출처: KD너겟, 2015년도 조사 자료

마. 고급 통계 분석 기법의 활용

인메모리 분석, 인-DB(In-DB) 분석 환경과 오픈소스 R의 대중화로 인하여 기존에 활용하기 어려웠던 다양한 분석 방법론들이 일반 사용자들에게 쉽게 다가갈 수 있는 여건이 형성되고 있다. 머신러닝과 관련하여 예측모델링 부분에서도 비모수적 검정과 함수 추정 방법을 활용한 정교한 모델링의 활용도가 높아지고 있다.

3. 기대 효과

데이터분석 컨설팅은 비즈니스에 실질적인 성과를 가져다 줄 수 있는 모형과 실행 가능한 의사결정 지원 시스템의 구축으로 연결된다. 과학적인 의사결정 과정이 실질적인 성과로 연결될 수 있음을 보여줄 수 있다면 관련 분야의 활성화는 더욱 가속화될 것이다.

데이터 분석 과정은 기존 구축 인프라의 활용과 직접적인 연관성을 가지고 있어 기 구축된 빅데이터 인프라와 다양한 데이터 관련 사업들을 통하여 구축된 기반 인프라를 활용할 수 있는 가장 좋은 활용 모델이다. 데이터 분석 컨설팅을 통하여 기존 구축 인프라의 활용성을 높일 수 있고, 관련 분야의 활성화도 기대할 수 있다. 공공기관 등에서 최근 빅데이터 플랫폼 구축 이후 데이터 분석 모델의 개발에 집중하는 이유도 여기에서 찾아볼 수 있다. 특히 머신러닝과 같이 기반 인프라의 활용도가 높고 고도화된 분석 컨설팅 영역이 실제 비즈니스에 긍정적인 효과를 가져오게 되면 관련 분야의 급속한 확대가 기대된다.

4. 향후 전망

가. 데이터

비정형 데이터의 집적과 활용도를 높일 수 있는 빅데이터 플랫폼의 발달, IT 환경의 고도화 및 머신러닝 등 분석 방법론의 확대로 더욱 다양한 데이터의 활용이 예상된다. 특히 체계화된 데이터 컨설팅을 기반으로 하는 빅데이터 플랫폼은 데이터 분석 과정에서 발생하는 지루한 데이터 사전 준비 과정을 단축시키고 비정형 데이터의 활용성을 더욱 높일 수 있어, 음성 기반의 데이터, 텍스트 기반 데이터 및 소셜 미디어 기반의 데이터 수요가 크게 증가할 것으로 보인다.

나. 인력

데이터 분석 컨설팅 영역은 각 업종별로 일정 수준의 시장 규모가 유지되고 있으나 관련 전문가들의 수급은 원활하지 못한 상황이다. 전문가들의 수급 문제가 분석 컨설팅 영역의 시장 규모 확대의 큰 걸림돌 중의 하나로 여겨지고 있었고, 그 상황은 현재도 크게 변하지 않고 있다. 향후 빅데이터 분석, 머신러닝의 활용도 증가 등에 따른 관련 인력 수요는 지속적으로 증가할 것으로 보인다.

제6절 비즈니스 인텔리전스 구축

1. 비즈니스 인텔리전스 개요

비즈니스 인텔리전스(이하 BI)란 1989년 최종 사용자가 구축된 시스템 정보에 접근하거나 분석하는 방법을 뜻하는 의미로 처음 사용되었다. BI는 비즈니스 의사결정을 위해 기업 내외부 데이터 흐름에 대한 전략적 접근을 통해 분석데이터를 선별하고 접근·수집·통합·분석·추론의 과정을 통해 비즈니스 정보를 원하는 형태로 가공하여 제공하는 솔루션을 의미한다. 기술의 진보로 인해 BI는 사용자 요구사항 구현 범위를 넓혀가고 있으며 이를 뒷받침해 주는 기반기술이 다양한 형태로 활용되고 있다.

전통적으로 BI는 DW를 기반으로 한 다차원 모델 중심으로 구축되어 왔다. 정보계 프로젝트 수행을 통해 정의된 주제 영역을 기반으로 사전 설계된 DW마트 형태의 데이터 구조에서 보고자 하는 관점의 정보들을 간단한 조작만으로 손쉽게 빠르게 획득하도록 구성하여 왔다. 이를 위해 DW, 올랩, ETL의 기반기술을 바탕으로 데이터 품질관리, 데이터마이닝, 데이터 모델링, 대시보드/리포팅의 요소기술들이 같이 발전하였다. 이에 따라 BI는 정확한 분석과 예측 능력을 통해 기업 경쟁력과 성장을 도모하기 위한 비즈니스 의사결정의 중요한 도구로 자리잡게 되었다.

최근 사용자 요구사항 및 데이터 증가에 따른 IT 기술의 발전으로 데이터 처리 기술은 비즈니스적인 관점과 기술적인 관점에서 계속 발전해 왔으며, 특히 빅데이터 기반 기술에 따른 데이터 아키텍처의 발전은 BI의 발전 가능성을 더욱 빠르게 진화시키고 있다.

기술의 진보로 인해 BI에서 얻어지는 가장 큰 장점은 대용량 데이터 처리를 위한 데이터 처리 속도 증가와 사용자가 원하는 데이터를 표현하는 데이터 가공기술의 발전을 들 수 있다.

가. DW 아키텍처의 변화

현재 병렬처리 기반의 DW 어플라이언스로 대표되는 DW 아키텍처 시장에선 메모리 가격하락과 데이터 처리기술의 발전으로 인-메모리 DB를 통해 사용자 동시접속과 응답속도를 높일 수 있는 실시간 데이터 분석 아키텍처 제품이 시장에 선을 보이기 시작하였다. 또한 하둡, 클라우드 등 대용량 데이터 처리기술을 이용한 빅데이터 아키텍처를 통해 정형·비정형 데이터를 분산처리 방식을 통해 빠르게 추출할 수 있는 빅데이터 분석 시스템이 도입되고 있다.

그림 3-3-11 BI 구성을 위한 기술 아키텍처



나. BI 환경의 변화

데이터의 증가와 데이터 가공기술의 발전은 BI가 특정 사용자가 아닌 모든 비즈니스 사용자가 자신에게 필요한 분석업무를 직접 수행하는 셀프 서비스 비즈니스 인텔리전스로 확대되어 이를 지원하는 차세대 의사결정 지원 도구를 통한 다양한 비즈니스 활동이 가능할 것으로 보고 있다.

차세대 의사결정 지원도구의 특징

- 대용량의 데이터에 대한 직관적인 분석을 위한 시각화 도구
- 모바일 도구를 통한 즉각적인 데이터 분석과 활용 확대
- 메모리 기반의 데이터처리를 통한 분석 데이터의 응답성 향상
- 배치 데이터 기반이 아닌 실시간 BI 등장(예. 스트림 분석)
- 정형 데이터/고정화된 모델이 아닌 비정형 데이터/가변데이터 모델 지원

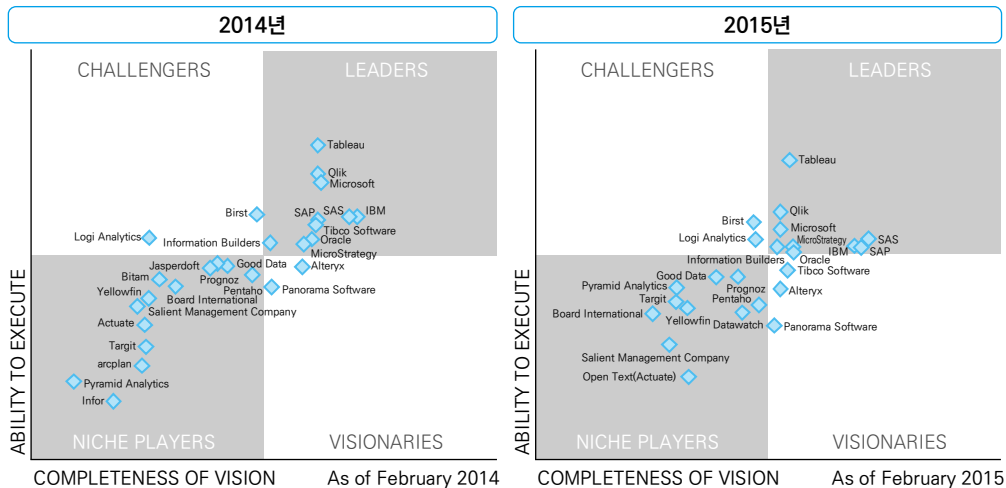
2. 비즈니스 인텔리전스 동향

지금까지 많은 기업들이 DW 기반의 BI를 구축하여 왔으며 축적된 데이터를 활용하여 비즈니스 경쟁력을 높이고자 지속적인 노력을 하고 있다. 최근 비즈니스 분석(BA, Business Analytics)이 주목받는 것도 기업의 경쟁력을 높이기 위해 비즈니스 사용자의 의사결정 수준을 높이려는 것과 무관하지 않다.

전통적인 BI와 BA의 차이에 대해서는 기업마다 다른 관점을 가지고 있다. 사용자 요구사항을 구현하는 프로세스적인 측면에서 BI와 BA를 구분하기도 하고 BI에 비해 직관적으로 의사결정에 도움을 줄 수 있는 비즈니스 인사이트 정보나 예측정보를 제공하는 도구를 BA라고 이야기하기도 한다.

BI가 데이터 분석기술에서 출발하여 비즈니스 문제를 해결하는 형태로 발전해 왔다면 BA는 비즈니스 문제해결이라는 측면에서 분석기술을 활용하는 목적으로 시작되었다. IT-Centric-Centralized와 Business-Centric-Decentralized으로 구분되고 있는데 이렇듯 시작의 차이는 존재하지만 BI와 BA가 지향하는 목표점은 비즈니스 문제해결 및 의사결정을 위한 데이터를 제공한다는 점에서 동일하며 아래 [그림 3-3-12]에서도 BI와 BA를 동일한 관점에서 비교하며 BI의 확장개념으로 보고 있다.

그림 3-3-12 BI & 분석 부문 매직 쿼드런트



※ 출처: The 2015 Magic Quadrant for Business Intelligence & Analytics, 가트너

3. 향후 전망

최근 국내에서도 빅데이터와 예측 분야에서 BA 도구를 적용하는 사례들이 늘고 있으며 IT 프로젝트에서 BI 도구와 BA 도구가 경쟁하는 모습을 많이 볼 수 있다. 그러나 앞에서도 언급된 것처럼 사용자 입장에서 BI와 BA를 정확하게 구분하여 도입할 필요는 없다.

전통적인 BI 기업들의 경우도 빅데이터 지원, 비정형모델 지원, 시각화, 인메모리 분석, 예측분석 등의 기능을 강화하고 있어 BI와 BA는 분석도구의 구분보다는 발전된 인텔리전스 & 분석(Intelligence & Analysis) 접근법으로 보는 견해도 많다.

향후 기업 활동에서 전통적인 방식의 분석 기법은 광범위한 사용자들이 업무에 직접적으로 활용할 수 있는 실시간 분석데이터 제공 형태로 발전될 것으로 보이며 리포팅과 대시보드 등 사용자 환경에서 직관적인 데이터 분석을 위한 시각화 도구의 사용도 늘어날 것으로 전망된다.

또한 프로젝트 중심의 DW/DM 구조에 종속적인 데이터 분석의 한계를 극복하기 위해 데이터 구조에 비종속적인 분석 모델의 구현, 비정형 데이터를 활용한 데이터분석, 개발자와 분석사용자 간의 애자일 방법론 등의 점진적인 구축 프로세스를 통한 전략적 데이터 분석환경을 구성하는 변화가 국내에도 시작될 것으로 보인다.

최근 가트너는 ‘2016년 정부를 위한 10대 전략적 기술 트렌드(The Top 10 Strategic Technology Trends for Government in 2016)’ 보고서를 통해 분석은 데이터를 수집 및 해석함으로써 조직적 효율성 또는 프로그램 효과를 높일 수 있는 통찰력을 제공한다고 말한다. 업무 활동 및 서비스 제공의 모든 단계에서 분석을 보편적으로 활용하는 이른바 ‘보편화된 분석(analytics everywhere)’으로 지표들이 나열된 대시보드 리포팅에서 탈피하여 실시간으로 보다 정확한 맥락 기반의 의사결정을 내릴 수 있도록 돕는 자율적인 비즈니스 프로세스 및 비즈니스 인텔리전스로 전환되고 있다는 것이다.

사용자가 운영 업무 중에 즉각적으로 의사결정에 필요한 데이터를 얻어 바로 업무에 적용하는 보편화된 분석의 이야기는 과거에도 많이 이야기되고 있었지만 데이터의 즉시성, 편의성의 측면에서 많은 제약이 있었던 것이 사실이다. 그러나 최근 데이터 처리 기술의 발전과 다양한 사용자 도구들은 과거의 한계를 넘어 비즈니스 요구사항을 현실화하는 데 한걸음 더 나아갈 것으로 전망된다.

제7절 오픈 데이터 구축

1. 오픈 데이터 구축 개요

가. 오픈 데이터

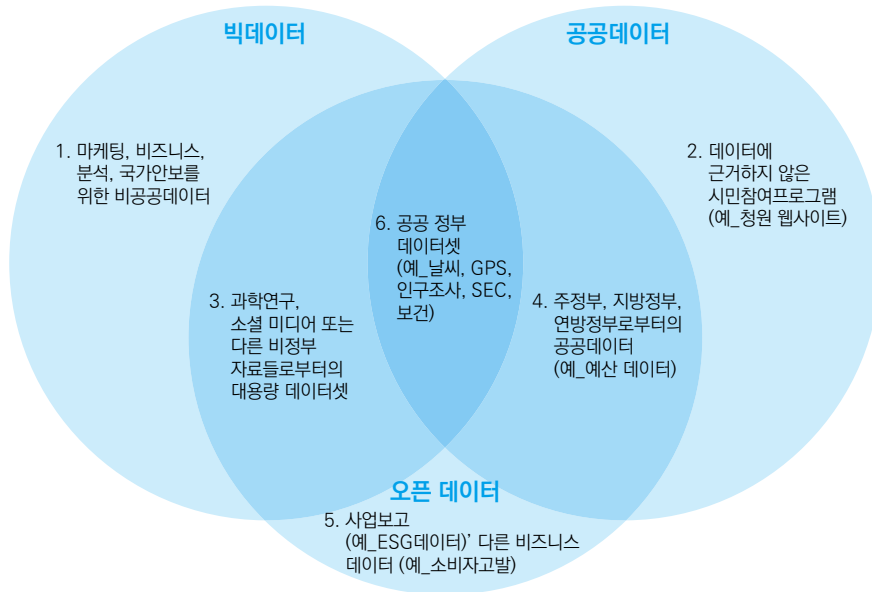
2015년 5월 당시 국내 대표 내비게이션 앱인 ‘김기사⁶¹⁾’가 카카오에 626억원에 인수되어 세상을 놀라게 하는 사건이 있었다. 김기사는 오픈 데이터를 통해 사업모델을 만들어 성공한 대표적인 케이스이다. 2012년 영국의 팀 버너스 리(Tim Berners-Lee)는 오픈 데이터 즉, 국가나 공공기관이 보유하고 있는 각종 데이터가 새로운 기회의 원천, 부가가치 창출의 우물이 될 것이라고 예언한 바 있다. 이에 화답하여 영국 정부는 선뜻 170여억원을 그들이 세운 오픈데이터연구소(ODI, Open Data Institute)에 지원했는데 국가적 차원에서 공공데이터의 가치를 인정하고 이것의 활용이 새로운 미래의 경쟁력이 될 것임을 알고 있었다는 것이다.

다른 나라들도 공공데이터 활용에 대해 다양한 정책을 추진하고 있다. 미국은 2012년 5월 투명성 제고, 일자리와 비즈니스 창출을 위해 모든 공공정보를 기계 판독 가능한 오픈포맷으로 개방하라는 ‘대통령 행정명령과 오픈 데이터 정책’을 발표하였다. 유럽연합도 2011년 12월 공공데이터를 온라인 단일 창구를 통해 무료로 제공해 경제적 기회 증대 및 사회현안 해결에 활용한다는 ‘오픈 데이터 전략’을 발표했으며 일본의 아베총리도 “전자행정 오픈 데이터 전략은 행정의 신뢰성·투명성 향상, 신시장 창출에 따른 경제 활성화 관점에서 매우 중요하기 때문에 공공데이터를 적극 개방하고 활용을 촉진하기 위한 구체적인 방안을 검토할 계획”이라고 밝혔다. 이러한 정책에 따라 세계 각국의 공공데이터 개방과 민간 활용이 가속화 되고 있는 상황에서 우리나라에서도 공공데이터 활용으로 경제에 새로운 활력소를 불어 넣고 성장의 동력으로 삼아야 할 시점이 이미 도래했다.

오픈 데이터는 저작권, 특허 또는 다른 통제 매커니즘 없이 특정한 데이터가 모든 사람에게 대가없이 제공해야 한다는 생각에서 출발하였다. 오픈 데이터 운동의 목표는 오픈소스, 오픈 하드웨어, 오픈 콘텐츠, 오픈 액세스 등의 ‘개방’ 운동과 유사하다. 오픈 데이터 이면에 숨겨져 있는 철학은 오랜 역사를 갖고 있지만 인터넷과 월드와이드웹의 등장으로 최근에 출현한 용어이다. 특히 미국과 영국의 ‘Data.gov’, ‘Data.gov.uk’ 등의 이니셔티브에 의해 촉발된 매우 뜨겁고 현재진행형인 이슈라고 할 수 있다.

61) 김기사 : 2016년 2월부터 ‘카카오내비’로 명칭 변경하여 서비스 중임

그림 3-3-13 데이터 카테고리 정의



※ 출처: "Open Data Now"(J. Gurin) 2014

오픈 데이터는 비즈니스, 소셜, 분석, 과학 등 민간 영역의 데이터와 정부가 업무상 생성하여 공개하는 공공데이터가 중첩되어 있는 포지션을 가지며 여기에 ESG(Environment, Social, Governance)⁶²⁾ 데이터를 포함한다. 오픈 데이터 엔터프라이즈 설립자 조엘 구린(Joel Gurin)은 그의 저서인 『오픈데이터나우Open Data Now』에서 “오픈 데이터는 사람, 기업, 조직이 새로운 사업을 시작하고 패턴과 트렌드를 분석하며 데이터 중심 결정을 내리고 복잡한 문제를 해결하기 위해 사용할 수 있는 접근 가능한 공공데이터로써 누구나 사용할 수 있도록 반드시 공개적으로 사용 가능해야만 하고 재사용을 허용하는 방식으로 라이선스가 부여되어야 한다”고 언급하면서 “오픈 데이터의 사용료는 무료 혹은 최소한의 비용이어야 한다는 일반적 합의가 존재한다”고 했다.

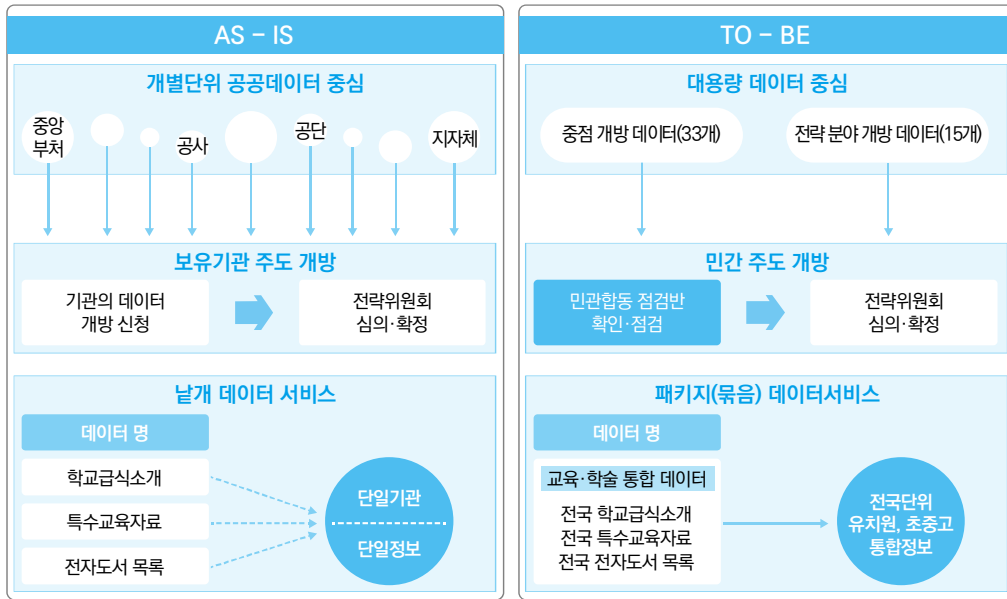
나. 오픈 데이터 구축 개요

현재 우리나라에서 이루어지는 오픈 데이터 구축은 위에서 언급한 데이터의 성격에 따라 데이터의 공급자로서 정부가 그 중심에서 주도하고 있다. 기존 오픈 데이터 구축·제공의 한계는 1)제공기관 중심의 데이터 제공 2) 품질 및 표준화 미흡에 따른 활용제한 3)제공기반 미흡으로 접근성 불편 등의 문제가 제기되어 2014년 하반기 「공공데이터 개방 발전전략」을 발표하고 산업적 파급

⁶²⁾ ESG(environment, social, governance): 기업의 사회·환경적 활동까지 고려하여 기업의 성과를 측정하는 기업성과지표

효과가 높은 대용량 공공데이터를 우선 개방하기로 하고 대용량 패키지 개방을 실시하고 있다.

그림 3-3-14 오픈 데이터 구축 모형



※ 출처: 공공데이터 개방 발전전략, 한국정보화진흥원, 2014

오픈 데이터 구축은 민간이 공공데이터를 원활히 활용하도록 데이터품질, 제공 포맷과 방법, 제공인프라 등을 향상시키는 데 그 목적이 있다. 이에 따라 현재 이루어지는 오픈 데이터 구축은 개방체계 구축, 데이터품질 진단 및 개선, 개방 DB 구축 및 오픈 API⁶³⁾ 개발, LOD⁶⁴⁾ 구축 등의 형태로 추진되고 있다. LOD 구축은 2014년 부터 한국수자원공사, 국립수목원, 국립중앙도서관 등이 개방 특성에 따라 추진되고는 있으나 데이터 개방이 파일다운로드 및 오픈 API 중심으로 이루어지면서 구축 규모는 크지 않다.

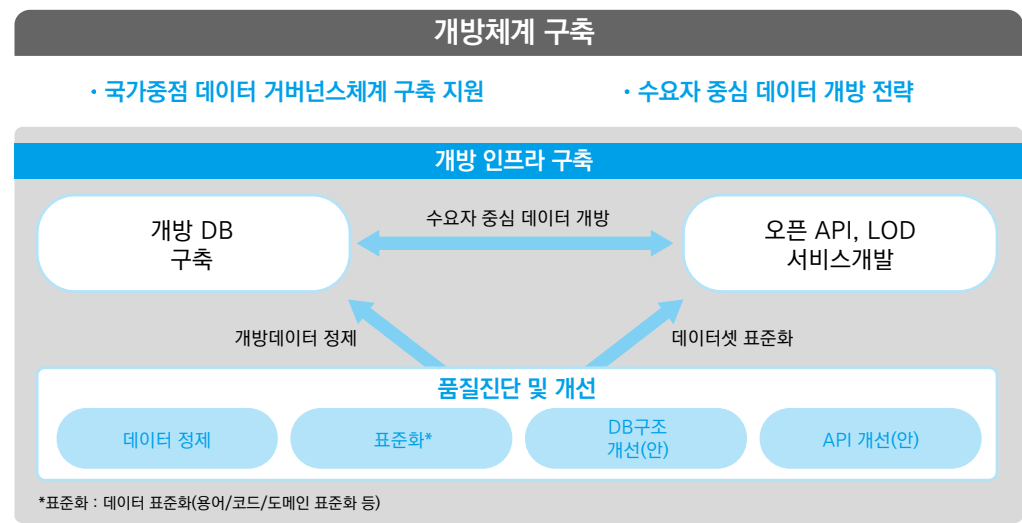
오픈 데이터 구축은 위에 언급한 5개의 형태가 현재 데이터빅뱅 프로젝트라는 이름으로 이루어지고 있다. 각 구축 형태별로 범주와 구축내용을 간략히 짚어보고자 한다.

⁶³⁾ 오픈 API: 외부에 개방된 API로서 데이터를 일정한 형식으로 일반에 개방하여 그 데이터를 이용하거나, 직접 응용 프로그램과 서비스를 개발할 수 있는 방식(기관의 데이터가 갱신되면 실시간으로 이용자가 제공받아 이용 가능한 방식)

⁶⁴⁾ LOD: 사용자가 정확하게 원하는 정보를 찾을 수 있도록 웹상의 모든 데이터와 데이터베이스를 공개하고 연결하는 것. HTTP, RDF(resource description framework)와 URI(Uniform Resource Identifier) 등의 웹 표준 기술을 이용하여 시맨틱 데이터를 구축하고, 출처가 서로 다르지만 URIs를 통해 데이터를 서로 연결함으로써 웹에 공개, 연계, 공유하는 기술



그림 3-3-15 오픈 데이터 구축 연관도



※ 출처: 데이터 빅뱅 프로젝트 수행과제 간의 상호연관 관계, 한국정보화진흥원, 2016

1) 개방체계 구축

개방 전략 수립은 대상 DB와 관련된 업무·정보화 환경을 분석하고 수요자 중심의 대용량 DB 개방·관리·활용 체계가 실현되기 위해 데이터 개방계획을 제시하고 지속적이고 발전적인 데이터 관리 체계 구축을 위한 조직·예산구성(안), 지침 및 가이드 제시, 변화 관리를 위한 관리프로세스 및 인프라 구성안을 제시한다. 이를 위해 수요자 관점의 니즈 파악을 위한 정보 수요자 인터뷰, 개방 시나리오 개발을 통한 관심이 높은 개방 데이터 발굴 등이 병행된다.

2) 품질진단 및 개선

개방 DB, 오픈 API 서비스, LOD 서비스 등의 개방 서비스로 정보 수요자가 보다 가치있게 데이터를 활용하기 위해서는 대용량 DB 자체의 데이터 품질이 확보되고 데이터 관리 체계가 내재화 되어야 한다.

품질진단 및 개선은 대용량 DB의 값·구조·표준·연계·품질관리 수준진단을 통해 데이터 품질개선 기회를 식별하고 해당 과제기간 내에 최대한 가능한 범위 내에서 데이터 정제, 데이터 표준화, 성능 튜닝, 응용기능(AP) 개선(혹은 개선가이드 제시) 등의 품질개선을 실시하고 프로젝트 이후에도 체계적인 데이터 품질관리를 통해 지속적이고 반복적인 품질개선이 가능하도록 방안을 제시한다.

3) 개방 DB 구축

대용량 DB를 개인정보 제거, 데이터 형식 및 포맷 표준화 등을 거쳐 개방 데이터를 자동으로 수집→가공→적재하는 형태로 별도 개방 DB를 구축하는 작업을 거쳐 정보수요자들이 원문을 직접 다운로드할 수 있게 하거나 오픈 API 서비스에 필요한 데이터를 DB로 구축하는 것이 주요 내용이다.

이를 위해 대용량 DB에서 개방범위를 식별하고, DB구축을 위해 분석→설계→메타데이터 구축→적용 단계를 거쳐 개방 DB를 구축한다. 그리고 경우에 따라서는 개방이 필요한 데이터를 확보하기 위해 원문자료의 디지털화, 데이터 재가공 등의 작업도 개방 DB 구축영역에 포함된다.

4) 오픈 API 개발

개방 데이터를 실시간으로 제공하기 위해 오픈 API를 개발하여 다수의 활용자가 오픈 API에 연계하는 방식으로 데이터를 활용할 수 있도록 서비스를 개발한다.

오픈 API는 일반적인 소프트웨어 개발 방법론과 동일하게 분석→설계→개발→시험 및 적용 과정을 거쳐 개발한다. 모든 오픈 API 서비스는 공공데이터 포털(data.go.kr)에 등록되어야 하며 대용량DB 보유기관이 별도의 개방 서비스 플랫폼을 구축할 경우 공공데이터 포털과 연계가 되어야 한다.

5) LOD 구축

개방하고자 하는 데이터(셋)를 웹을 기반으로 누구나 자유롭게 사용, 재사용, 배포가 가능하도록 URI(Uniform Resource Identifier), RDF(Resource Description Framework), SPARQL(Simple Protocol And RDF Query Language) 등 기술 요소와 표준 용어집 등을 활용하여 상호 운용성이 보장되는 기계 가독형 데이터셋으로 표현해야 한다. 기관이 보유한 데이터 및 국내·외의 연계 가능한 데이터를 조사하여 LOD 구축 대상 데이터를 선정, 원천 데이터를 확보하고, 시범데이터 구축을 통해 실 구축 중 발생할 수 있는 문제점 및 해결방안을 마련한다. 구축 및 발행단계에서는 원천 데이터를 분석하여 표준화된 용어를 정의하고, 도메인 모델을 구성하는 클래스, 속성 및 관계 정의를 통해 온톨로지⁶⁵⁾를 작성, 데이터의 정제 및 표준화된 방식으로 변환하여 제공한다.

오픈 데이터 구축 사업의 예산 규모는 2013년부터 연간 120여 억원 규모로 꾸준히 이루어지고 있다[표 3-3-4]. 하지만 2014년 하반기에 발표한 공공데이터 개방 발전전략에 따라 2015년부터는 기존에 기관별로 각각 데이터 품질지원, 오픈 API 구축, LOD 구축 등 개별적으로 지원하는 방식에서 공공데이터 활용의 시너지를 높이기 위해 단일 기관에 대해 요건이 있을 때 종합지원 방식(데이터빅뱅)으로 추진되고 있고 규모도 약 70여 억원을 상회하는 수준이다.

⁶⁵⁾ 온톨로지(ontology): 합의된 지식을 컴퓨터에서 다룰 수 있는 형태로 표현한 모델로, 개념의 타입이나 사용상의 제약조건을 명시적으로 정의한 기술

표 3-3-4 연도별 오픈 데이터 구축사업 예산 현황

(단위 : 억원)

사업구분	2013년	2014년	2015년	2016년
품질관리	60	40	8.7	23.6
오픈 API	43	48.7	27	24.5
LOD	-	6.5	6	-
데이터빅뱅	-	-	72	75
DB구축	15.3	20.2	5	6
전체	118.3	115.4	118.7	129.1

※ 출처: 공공데이터 이용 활성화 지원 사업, 조달청나라장터·NIA, 2016.

※ 기재된 예산규모는 NIA자료 및 나라장터의 발주기준으로 정리된 것으로 계약금액과는 다름.

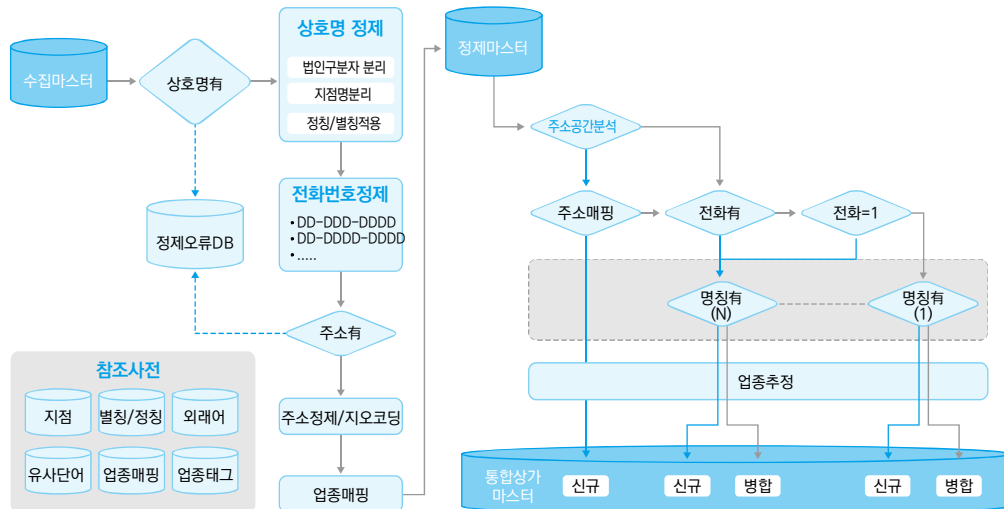
2. 오픈 데이터 구축 동향

가. 소상공인시장진흥공단 상권정보 마스터구축 및 API 서비스

소상공인시장진흥공단은 기존에 전국 주요 상권 현황, 유동인구 정보 2종을 다운로드 형태로 개방 중이었으나 상가데이터의 부정확성 및 정보 왜곡에 따른 과잉 창업 폐해 발생으로 인한 사회적 비용이 발생하였다. 이에 따라 전국 단위 고품질 상가정보 개방을 위해 수요자 중심의 개방 전략을 수립하고, 상가정보 활용성 증대를 위해 시범지역을 선정하여 품질 진단 및 심화개선을 실시하고, 상가정보의 공동 활용을 위해 상가마스터DB 구축 및 오픈 API를 개발하였다. 이 구축사업으로 상권DB의 중복오류 정제 및 최신화를 통해 개방 데이터의 신뢰성 확보 및 민간 활용성을 제고했으며 상권정보의 데이터 표준화를 통한 연계기관 간 정보 일관성 확보 및 오류로 인한 재처리 비용을 절감하게 되었다. 아울러 관련 데이터를 연계한 빅데이터 분석 및 매시업 서비스 확대를 통한 신규 서비스 발굴, 신사업 창출 등의 효과도 기대된다.

또한 상권DB의 품질향상으로 인한 상권 분석의 정확도 개선으로 예비창업자에게 정확한 정보 제공, 과잉창업 예방 및 폐업을 감소로 인해 자영업 실패방지로 인한 연간 2,300억원의 사회적 비용절감 효과가 있을 것으로 보여진다.

그림 3-3-16 상가정보 통합엔진을 통한 개방데이터 구축

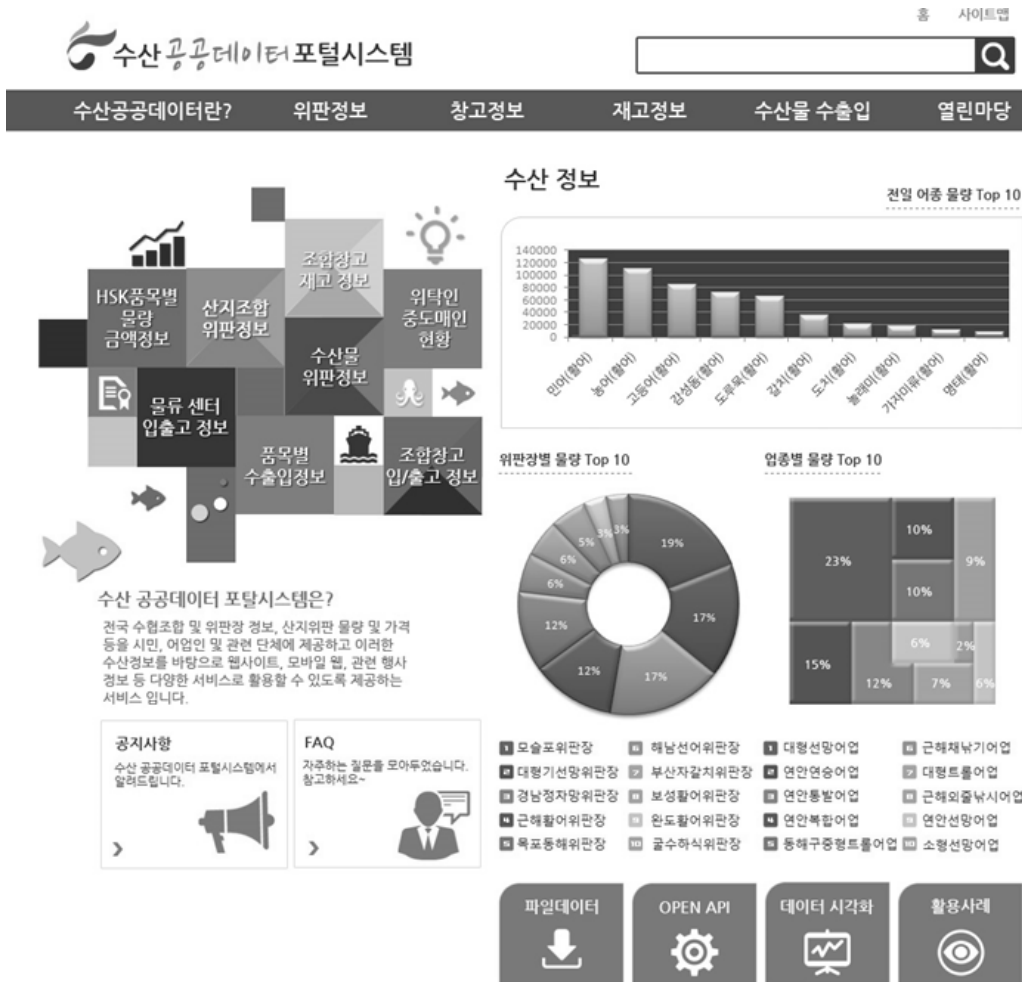


※ 출처: 상권정보 마스터구축 및 API서비스, 소상공인시장진흥공단, 2016

나. 해양수산부 수산정보 개방 DB 및 API 서비스

수산정보는 수산정보포털(<http://www.fips.go.kr>)을 통해 수산 관련 정보들을 전체 데이터가 아닌 집계된 데이터로 제공 중이다. 수산정보포털에서 엑셀 다운로드 형태로 제공되며 실시간으로는 제공되고 있지 않아 민간에서 활용도가 낮고 조회 기간의 제한으로 다량의 데이터 또는 월 단위 데이터를 한꺼번에 조회하기 어려웠다. 이에 따라 해양수산부는 개방되는 위판정보, 수산물 수출입 정보에 대한 값·구조·연계·표준 관점의 데이터 품질 진단 및 개선을 추진하고 농림축산식품부의 품목분류코드, 해양수산부 수산물 표준코드 등 유사한 코드를 상이하게 관리함에 따른 불합리함을 해소하기 위해 코드 통합 및 신규 수산물 표준코드를 정의하였다. 그리고 수산정보 개방을 위한 기업, 위판장, 중도매인, 어민을 대상으로 수요자 인터뷰를 실시하여 어민의 경우 일자별 위판되는 어종의 위판가격 변동정보가 위탁판매장 선택의 중요한 의사결정 요소이며, 위판장을 쉽게 찾을 수 있는 주소 정보도 활용도가 높은 정보를 식별하였다. 수요자 인터뷰/설문, 요구사항 도출, 선진사례 분석결과를 바탕으로 수산정보DB의 연계누락, 오류데이터 정비 등 데이터 품질진단 결과를 바탕으로 개방DB를 구축하여 7종의 고수요·고가치 수산정보 개방 데이터셋을 정의하여 오픈 API 서비스를 개발하였다. 이 사업의 결과로 위판장별, 어종별 위판량 및 가격정보, 재고정보, 수산물 수입정보를 제공함으로써 연간 약 36억원의 어업인 소득증대가 기대되며, 아울러 수산물 유통 및 수출입 정보 개방을 통한 수산물 정보제공 서비스 개발로 부가가치 및 일자리 창출과 수산업 관련 종사자인 수출입 유통기업, 수산물 소비자에 대한 소득증대 및 가계비용 절감 등이 기대된다.

그림 3-3-17 수산정보 개방포털 구축



※ 출처: 수산정보DB구축 및 API서비스, 해양수산부, 2015

다. 산림청 등산로 개방 DB 및 API 서비스

산림청 ‘숲에온’은 2011년부터 숲길기본계획에 따라 국민의 등산, 트래킹, 산악레저스포츠, 탐방 및 휴양·치유 등의 활동 증진을 목적으로 등산로 DB를 구축하여 ‘숲에온’을 통해 서비스하고 있다. 휴양정보의 기초정보인 등산로의 경우 산림청, 국립공원관리공단, 일부 지자체 등을 통해 분산하여 개방·관리됨에 따라 이를 활용하는 기업과 민간에서는 제공받은 데이터의 표준, 품질기준 및 서비스 제공 창구가 서로 달라 활용에 제약이 많았다.

등산인구의 증가에 따라 개방된 자료는 대국민 또는 관련 활용기업에 반드시 필요하고 수요 또한 지속적으로 증가하고 있다. 수요조사 결과에 따르면 등산로 노선 정보의 경우 신규 및 잠재 수요자

는 대형포털업체와 등산로정보·국가생물종정보·지역정보 등을 연계하여 새로운 서비스를 기획하고 있는 레저기업 등에서 새로운 비즈니스 창출의 기회로 활용을 기대하는 상황이었다. 이에 따라 산림청은 오픈 데이터 구축사업에 등산로 및 국가생물종DB 데이터 품질 개선을 진행하고 산 정보 3,368개 정비와 등산로 노선 9,000km를 현행화하고 오픈 API로 구축하였다. 그리고 민간 활용 확대를 위해 산림공공데이터 활용 경진대회를 개최하였다. 이 사업의 결과로 산정보, 등산로, 숲길, 숲해설 등의 ‘숲에온’ DB와 문화관광 DB를 연계·융합하여 해당 지역 방문을 유도, 지역경제 활성화를 도모하였다. 경북대 조사결과 지리산둘레길 인근지역의 경제적 파급효과는 약 352억원, 정보 개방을 활용한 신규 앱 개발 등 신규 산업 가치 창출은 연간 약 14억원 정도로 추정하였다.

그림 3-3-18 산림 공공데이터 포레스톤



※ 출처: 등산로개방DB 구축 서비스, 산림청, 2015

3. 오픈 데이터 구축 성과

가. 데이터 개방 현황

2013년 이후 지속적인 오픈 데이터 구축으로 국내외 개방 관련 지수가 크게 향상되었다. 2016년 5월 공공데이터 개방현황을 대표하는 공공데이터포털(www.data.go.kr)의 통계를 보면 중앙부처를 포함한 공공기관 중 656개 기관이 파일데이터 1만 4,771종과 오픈 API 1,953종을 개방하고 있다. 정부가 공공데이터 개방사업을 주도하면서 4년간 1만 7,000여 건을 공개했으나 민간에서 이를 활용해 개발한 사업은 853건으로 활용이 저조하다는 지적이 제기되고 있지만 공공과 민간은 서로 상생의 생태계를 만들어야 하는 공동책임이 있는 만큼 앞으로 지속적으로 개선해야 할 것으로 보인다.

표 3-3-5 공공데이터 개방현황

〈공공데이터 개방현황〉				
구 분	2013년	2014년	2015년	2016년 (5월 말 누적)
	누적결산 (2013년말)	누적결산 (2014년말)	누적결산 (2015년말)	
파일데이터	4,718	11,796	14,122	14,771
오픈 API	554	1,361	1,790	1,953
합 계	5,272	13,157	15,912	16,724

〈공공데이터 개방현황〉					
구 분	2013년	2014년	2015년	2016년	총계
개발 사례	42개	353개	300개	158개	853개

※ 출처: 공공데이터개방현황, 한국정보화진흥원, 2016

공공데이터포털에서 제공되는 오픈 형태별 활용 신청 상위 10개는 [표 3-3-6]과 같다. 활용도가 높은 오픈 API는 서민생활과 관련된 기상 정보, 대기오염 정보, 아파트매매가 정보 등이 상위를 차지하고 있고, 대량으로 처리하기 용이한 파일데이터는 창업과 밀접한 상가정보, 기업정보 등이 대부분을 차지하고 있다. 이렇듯 공공데이터는 날이 갈수록 그 활용사례가 늘어가고 있는 상황이다.

표 3-3-6 공공데이터 활용신청 현황(2016.6)

■ 오픈 API 활용신청 TOP 10

구분	기관명	파일데이터 명	확장자
1	소상공인시장진흥공단	상기업소DB	CSV
2	소상공인시장진흥공단	전국주요상권현황	CSV
3	중소기업청	전국 1,200대 주요상권DB	CSV
4	중소기업청	기업정보DB	XLS
5	서울특별시	서울특별시 지점별 유동인구 수	XLSX
6	전라북도 정읍시	전라북도 정읍시 CCTV	CSV
7	중소기업청	중소기업확인서를 발급받은 중소기업	CSV
8	중소기업청	중소기업확인DB	XLS
9	교육부	2015년 교육기본통계(대학)	XLS
10	소상공인시장진흥공단	상권정보업종 코드집	CSV

■ 파일데이터 활용신청 TOP 10

구분	기관명	오픈 API 명	신청 수
1	기상청	(신)동네예보정보 조회서비스	232
2	한국환경공단	대기오염정보 조회 서비스	194
3	국토교통부	아파트매매 실거래자료	98
4	한국관광공사	국문 관광정보 서비스	81
5	국토교통부	버스노선정보 조회서비스	75
6	국토교통부	아파트 전월세 자료	60
7	한국산업안전보건공단	물질안전보건자료	59
8	국토교통부	버스정류소정보 조회서비스	57
9	한국예탁결제원	주식정보 서비스	55
10	우정사업본부	도로명주소 우편번호 조회서비스	52

※ 출처: 공공데이터 활용신청 현황, 공공데이터포털, 2016.6

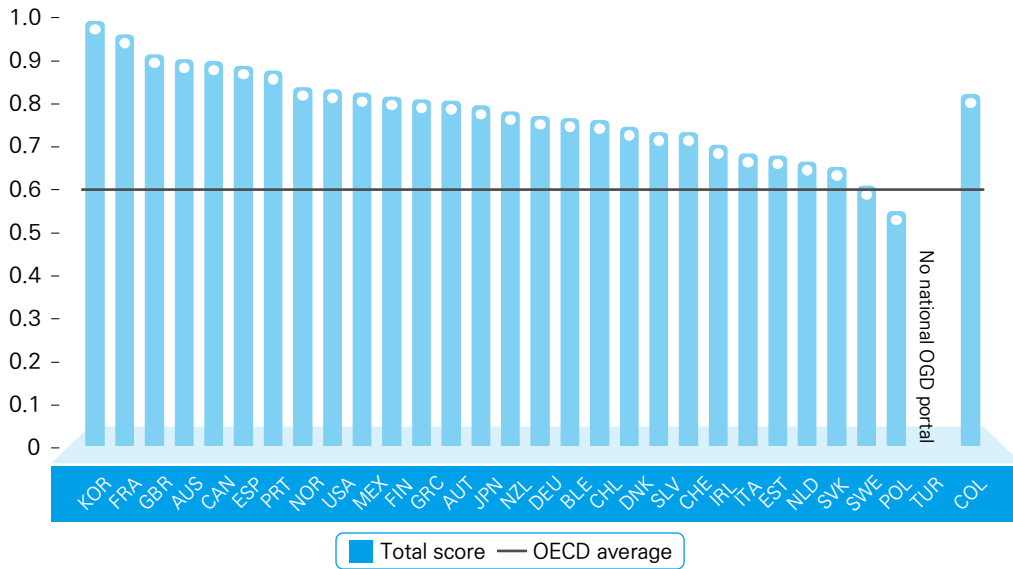
나. 데이터 개방 해외수준

2015년 7월 6일 OECD가 발표한 ‘2015년 정부백서(Government at a Glance 2015)’에서 우리나라는 공공데이터 개방지수(OUR⁶⁶⁾)에서 1점 만점에서 0.98을 달성해 여러 선진국을 제치고 1위에 올랐다. 오픈 데이터의 대표적 선진국인 영국과 미국이 각각 3위, 9위를 차지한 것을 보면 대단한 성과라고 할 수 있다. 아시아권 국가 중 일본은 14위를 기록해 OECD 평균을 약간 상회하

⁶⁶⁾ OUR(Open, Useful, Re-Usable): 공공데이터 개방지수는 OECD에서 각국의 공공데이터 개방 전략 수립 및 이행을 돕기 위해 개발한 것으로 가용성, 접근성, 정부지원 등 3개 분야 19개 지표로 구성되어 있다.

는 데 그친 것으로 나왔다. 우리나라는 3개 분야에서 모두 높은 수준을 기록했는데 가용성 분야에서는 공공데이터 활용 서비스가 407개로 가장 많은 점이 높은 평가를 얻었고, 접근성 분야에서는 공공데이터포털을 중심으로 민간에서 이용하기 쉽게 데이터를 개방하고 있다는 평가를 얻었다. 그리고 정부지원 분야에서는 정부와 공공기관 평가에 데이터 개방 항목을 반영한 점과 공공데이터를 활용한 창업을 적극 활성화한 점이 좋은 평가를 받은 것으로 보인다.

그림 3-3-19 2014년 OECD 국가 오픈 데이터 개방지수



※ 출처: 오픈 데이터 개방수준, OECD, 2015

4. 고려 사항

전세계적으로 오픈 데이터가 확산됨으로써 국가 전체 및 지역에서 의료보건, 농업, 기상, 교통, 보안, 환경 등 분야의 공공 서비스가 변화하고 있으며, 민간에서도 이를 활용한 스타트업 성공사례가 속출하고 있다. 공공부문에서 업무를 통해 만들어낸 수많은 데이터를 민간기업과 시민이 접근·이용·재사용이 가능하게 함으로써 정부의 신뢰성과 투명성 향상, 새로운 창업기회 창출, 국민 참여 및 의사결정에 많은 도움을 주고 있다는 것이다. 최근 4년여 간 정부는 오픈 데이터 구축을 적극적으로 추진함으로써 외형적으로 괄목할 만한 성과를 보여주고 있지만 선진국의 사례를 교훈 삼아 실제적인 효과를 누릴 수 있도록 다음의 사항을 고려해야 한다.



첫째, 성공적인 오픈 데이터의 시작에 이어 제도약을 위한 추진동력 확보가 필요하다. 현재의 오픈 데이터 정책은 정부의 창조경제 정책을 견인하기 위한 방안으로 설정한 좋은 선택으로 향후에도 지속적으로 우리 경제에 활력을 제공하는 역할을 하는 것이 중요하다. 따라서 일시적인 정책의 용도가 아니라 진정한 경제의 활력으로서 자리매김을 하려면 지금까지 진행한 오픈데이터의 정책, 사업, 법제도, 조직 등을 되돌아보고 새롭게 정비하는 것이 필요하다.

둘째, 좋은 활용사례에 대한 발굴과 홍보가 중요하다. 정부는 오픈 데이터의 제공자로서 활용 분야에는 역할상 한계가 존재하므로 활용과 활성화는 민간의 몫이다. 민간이 자체적으로 사례를 만들어 가고 있으나 그 속도를 빠르게 하기 위한 정부의 노력이 필요하다. 즉, 마중물의 역할이 필요하다. 민간이 파악하는 오픈 데이터는 자기가 알고 있는 영역과 그리고 인접한 영역의 데이터 들일 뿐이다. 활용 기반에 대한 효율적인 지원체계를 만들고 ‘우수사례’를 발굴하여 부지런히 알리고 자극해야 한다.

오픈 데이터는 산업의 새로운 자원이고 미래다. 2012년도 카이스트가 발표한 자료에 의하면 공공 데이터 활용가치는 연간 24조원, 15만명의 고용효과가 있다고 하였다. 유럽연합도 공공데이터 개방의 직접 효과는 27개국 전체를 대상으로 하였을 때 연 400억유로(약 60조원)의 가치가 있다고 하였다. 이처럼 공공데이터는 경제적으로 새로운 동력을 제공할 수 있는 21세기의 원유라고 볼 수 있다. 이 기회를 놓치지 않도록 민관이 서로 손잡고 가야 소기의 목적을 달성할 수 있을 것이다.



제4장 데이터 산업기술 동향

제1절 지능정보 기술

1. 개요

지난 수십 년간 지속되어온 정보화사회에서 한 가지 주목할 점은 개인용 컴퓨터(PC), 사물인터넷(IoT), 스마트폰 등과 같은 수많은 디지털 장비를 통해 엄청난 양의 데이터가 축적되고 있다는 사실이다. 여기에 컴퓨팅 파워의 향상, 기계학습(딥러닝) 기법의 발전 등이 더해지면서 인공지능(AI) 기술은 헬스케어, 지식서비스, 로봇, 금융, 교육, 엔터테인먼트 등의 분야에서 기존 산업의 외해적, 혁신적 기술발전을 견인하는 돌파구이자 고부가가치화를 주도하는 모멘텀으로 등장하였다.

지능정보기술은 ICBM 즉, 사물인터넷, 클라우드, 빅데이터, 모바일 서비스 등의 정보기술에 인공지능을 접목하여 신산업 혁신을 도모하고자 하는 키워드로 등장하였으며, 제4차 산업혁명으로 간주되는 ‘지식노동의 자동화’를 통해 지능정보사회로 진입하기 위한 핵심기술이다. 따라서 최근 많은 주목을 받고 있는 인공지능 관점에서 지능정보기술의 산업동향을 살펴보고자 한다⁶⁷⁾.

인공지능 기술은 구글, IBM, 애플, 페이스북 등 미국 중심의 글로벌 기업이 주도하고 있으며, 미국 외에도 유럽, 일본, 중국 등에서는 인공지능기술 확보를 국가적 아젠다로 채택하여 국가적 역량을 집중하고 있다. 반면, 국내에서는 단기간에 가시화할 수 있고 수요가 많은 분야 위주로 연구가 수행되고 있다. 장기간 연구수행이 필요한 분야는 연구 진행이 더딘 것으로 파악되고 있지만, 최근 AI연구소 설립 등을 포함한 국가 차원의 인공지능기술 육성에 대한 투자는 비록 선진국에 비해서는 늦었지만 향후 국내 산업혁신에 긍정적인 요소로 작용할 것으로 전망된다.

67) 국내외 산업동향은 한국산업기술진흥원(KIAT)의 산업전문인력역량강화사업 기획보고서(2016년 5월)의 내용을 바탕으로 재구성

2. 해외 현황

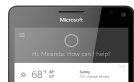
가. 미국

세계 인공지능 관련 기술을 주도하고 있는 미국에서는 구글, IBM, 애플, 페이스북 등 글로벌 기업이 무서운 속도로 성장하고 있을 뿐 아니라, 스타트업도 튼튼한 성장 기반을 이루고 있다. 이들 글로벌 기업은 이미 지능정보기술 개발에 경쟁적으로 투자를 확대하고 가시적인 성과를 도출하고 있다.

구글은 영국 소재 자회사인 딥마인드가 개발한 인공지능 바둑 프로그램인 ‘알파고’로 2016년 우리나라를 포함한 전세계에 큰 이목을 이끌었다. 이밖에도 방대한 데이터를 분석해 미래를 예측하는 기술인 머신러닝 프로그램 ‘텐서플로’를 공개했으며, 현재 자율주행자동차 개발을 완료하고 일반도로에서 시범 운행 중에 있다. IBM은 인공지능 컴퓨터 ‘왓슨’을 중심으로 눈에 띄는 움직임을 이어가고 있다. 왓슨은 2011년 미국 퀴즈쇼에서 인간을 상대로 압도적 우승을 거둔 인지 컴퓨팅 시스템으로 인간의 자연어 처리는 물론 대규모 데이터 분석과 학습이 가능하도록 구축되었다. 페이스북은 사람 얼굴 사진을 보고 누구인지 알아보는 ‘딥 페이스’, 페이스북 메신저에서 사용할 수 있는 가상비서 서비스 ‘M’을 출시하는 등 인공지능 기술을 접목한 서비스를 지속적으로 출시하고 있다. 애플도 2011년 개인비서 시리즈인 ‘시리’를 출시하며 인공지능 개발 흐름에 동참하고 있다.

표 3-4-1 인공지능 기술분야 미국내 글로벌기업 현황

회사명	주요 결과물	주요현황
구글	 AlphaGo	• AI기업 딥마인드를 약 5억 8,200만달러에 인수(2014.5) • 바둑프로그램 알파고 개발 및 이세돌 9단과 대국(2016.3)
	 TensorFlow™	• AI 기계학습 엔진 텐서플로의 오픈소스 공개(2015.11)
	 Introducing Google now for iPhone and iPad	• 검색엔진 기반 자연언어 처리기술 Now 개발 • 사용자의 사용패턴을 인지하고 명령전 정보를 제공
IBM	 IBM Watson	• 로봇, 의료, 헬스케어, 금융, 법률 등 응용 생태계 조성중(※ 36개국, 17개 산업 270여 개국 도입) • 소프트뱅크와 협력을 통한 로봇용 AI SDK 공개 예정
	 IBM Chef Watson	• 인지컴퓨팅을 기반으로 세상에 없는 요리법을 만들어주는 AI 요리앱 ‘셰프 왓슨’ 공개(2015.7)
	 IBM Watson Health	• 뉴욕암센터, MD앤더슨 암센터와 협업하여 진단 및 치료법 제안, 보험사 웰포인트는 치료적절성에 왓슨 권장
애플		• 음성인식기반 개인비서 서비스 ‘시리’를 통한 자연어 대화 기술을 스마트폰에 최초 탑재 • 스웰, 톱시, 보컬Q, 퍼셉티코, 이모션트 등 다수의 AI 스타트업 기업 인수

회사명	주요 결과물	주요현황
마이크로소프트		• 음성인식기반 가상비서 '코타나' 개발 • 코타나를 통해 컴퓨터, 스마트폰 등에 작업지시, 길 검색, 음식 주문 등도 가능, 키보드, 마우스, 화면터치 없이 음성만으로 점차 가능하도록 확장할 예정
페이스북		• 기계학습 기반 서비스지원이 가능한 AI 비서 M을 통해 온라인 쇼핑, 레스토랑 예약 등 작업 가능 • 메신저 연계 대화 가능한 AI 채팅봇 '챗봇' 공개 • 모노이덱스, 지비딕 등 다수의 AI 기업 인수

또한 미국은 정부차원에서 인공지능 원천기술 확보 및 군사 분야 접목을 통한 상용화 추진을 위해 2013년부터 브레인 이니셔티브(BRAIN Initiative) 사업을 추진하고 있다. 여기에는 7개 정부기관과 연구소, 대학 등 25개 민간기관이 참여하며, 향후 10년간 30억달러를 투자한다. 또한, 미국의 국방고등연구계획국(DARPA)은 인간 개입을 최소화한 무인드론 프로젝트(CODE 프로젝트) 및 인간의 뇌구조와 유사한 데이터 처리 칩셋 개발(SyNAPS 프로그램)을 2008년부터 진행하고 있다. 한편 스탠포드, MIT 등에서는 대학 내 AI 연구소를 통한 무인자율 주행차, 지능형 로봇 등의 분야에 다양한 AI 연구가 진행되고 있다.

나. 일본

지능정보 산업에서 일본 기업이 글로벌 주도권을 잡을 가능성은 매우 낮을 것으로 예상된다. 대다수 일본의 AI 활용 기업 및 로봇 기업들은 여전히 AI 기술로 구글, IBM 등이 개발한 솔루션을 채택하고 있는 실정이다. 그 동안 활용 가능한 수준의 자체 AI 기술 개발이 없었기 때문이다. 소프트뱅크의 휴머노이드 로봇 페퍼(Pepper)가 AI 기술로 IBM의 왓슨을 채택했고, AI를 활용해 마케팅 및 경영효율화 사업을 수행하는 다수의 중견기업들도 구글의 딥러닝 기술을 쓰고 있다. 다만 일본의 인공지능 산업에서 눈 여겨 봐야 할 것은 실제 산업 분야에서 AI 솔루션을 활용하려는 요구가 빠르게 증가하고 있다는 점이다. 또한 미국 기업 주도의 기술 방식과는 달리 자체 기술을 개발해 사업을 영위하려는 기업들이 최근 부각되고 있다.⁶⁸⁾

68) 출처: 4차 산업혁명 AI 기업분석, 한국투자증권 (2016. 4)

표 3-4-2 일본의 인공지능 대표기업 현황

회사명	주요서비스	주요현황
소프트뱅크		• 개인, 전문영역에서 활용 가능한 AI 로봇 '페퍼' 출시 • IBM 왓슨 기반 자연어 대화 및 맞춤형 서비스 진행 • MS 빅데이터 분석시스템 애저(Azure)를 통한 소매점 서비스, 관광안내, 노인 돌봄, 의료, 교육 등 AI 서비스 강화
		• IBM 왓슨 기반의 나옌 개발 • 미국 힐튼 호텔에서 투숙객 안내, 주변 관광 명소 추천 등 시범서비스
도요타		• AI를 총괄하는 Toyota Research Institute 설립, 5년간 약 10억달러(약 1.2조원) 투자하며, 로봇 및 자동차용 AI 연구
히타치		• 생산성 강화 및 로봇실용화를 위해 센서 및 AI 연구개발 분야에 약 2,000명 투입하여 물류용 AI 프로그램, 양팔로봇기술 등을 공개

일본 정부에서는 미래경쟁력 강화를 위해 로봇과 인공지능 분야를 집중 육성하고 있다. 일본재흥전략, 로봇신전략(Japan's Robot Strategy) 등을 발표하며 로봇혁명 핵심요소에 인공지능을 포함하였다. 이를 위해 2015년 5월에 경제산업성 산하에 AI 기술개발 및 실용화를 위한 AI 연구센터를 설립하였으며, 2016년 3월에는 총리 직속의 '인공지능 미래사회 경제전략본부'를 발족하였다. 또한, 경제산업성, 문부과학성, 총무성은 2016년부터 5개 산하기관에 10년간 1,000억엔을 투입해 AI 역량을 강화하고 로봇개발 등에 활용할 예정이다.

다. 중국

중국에서는 중국을 대표하는 IT 기업인 바이두가 인공지능 연구와 서비스 적용에 박차를 가하고 있다. 바이두는 2014년 머신러닝 전문가인 앤드류 응 스탠퍼드대 교수를 영입하면서 단숨에 인공지능 강자로 떠올랐으며, 중국 최대 검색 엔진을 보유한 강점을 살려 빅데이터를 기반으로 다양한 시도를 진행하고 있다. 바이두가 인공지능 기술 활용을 위해 노력을 기울이는 대표적인 분야는 무인자율주행 자동차다. 이를 위해 BMW와 협력하고, 2015년 12월 베이징 고속도로와 시내에서 자율주행 시범운행을 진행한 바 있다. 또한 중국의 대표적인 전자상거래 개발업체인 알리바바는 금융 결제, 전자상거래에 의한 데이터를 기반으로 인공지능 연구개발을 추진하고 있다. 인터넷서비스 및 게임서비스 전문기업인 텐센트 역시 인공지능 기사작성 로봇인 '드림라이터'를 개발하는 등 지능정보 기술 분야에서 중국 기업들의 글로벌 성장세가 빠르게 진행되고 있다.

표 3-4-3 중국의 인공지능 대표기업 현황

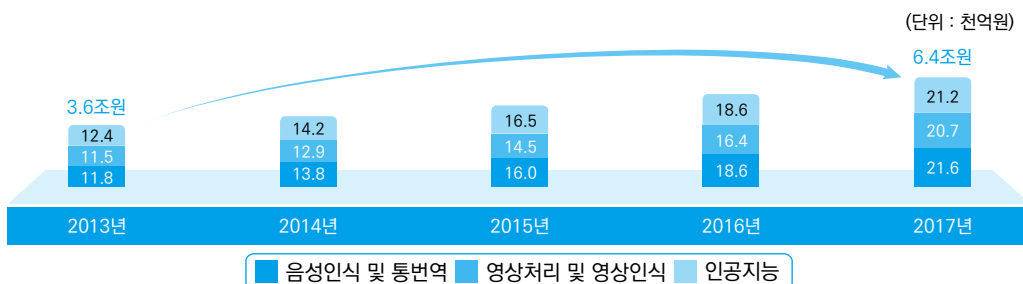
회사명	주요제품	주요현황
바이두	 비서로봇 두미	- 실리콘밸리에 딥러닝을 연구하는 연구기관 IDL(Institute of Deep Learning)을 설립(2013) - 구글 브레인 프로젝트 개발자 앤드류 응 교수 영입(2014) - AI기반 가상 비서로봇 '두미(度秘)'를 공개(딥러닝 융합 이미지, 음성인식 기술)
알리바바	 DT PAI	중국 과학기술부와 양자컴퓨터 전문 실험실을 공동 설립, 구글 알파고와 유사한 AI 개발 진행, 중국 최초 AI 플랫폼 'DT PAI'를 공개(2015)
텐센트	 드림라이터	스마트컴퓨팅검색실험실(TICS LAB)을 설립하여 AI 연구에 주력, 자체 개발한 AI를 탑재한 기사작성 로봇 '드림라이터' 출시(2015. 9)

중국 정부에서도 AI 육성 기반조성 및 전략적 투자 활성화를 위해 노력하고 있다. 중국의 제13차 5개년 계획(2016~2020)에 인간과 로봇의 상호작용 및 AI 관련 연구계획을 포함하였고, 중국 발전개혁위는 '인터넷+AI 행동실시 방안' 편성을 위한 TF 팀을 구성하고 AI 연구프로젝트인 '차이나브레인' 등 국가 차원의 실천방안 및 로드맵을 수립하여 발표할 예정이다. 또한 100대 국가전략사업 중 '뇌 과학과 두뇌 관련 연구'를 우선적으로 추진할 예정이다.

3. 국내 현황

국내의 인공지능 시장규모는 2013년 기준으로 약 3.6조원 정도(240조원 규모로 추정되는 세계 시장의 1.5%)에 불과하지만, 2017년에는 약 6.4조원 규모로 매년 15% 이상 성장할 것으로 예상되고 있다.

그림 3-4-1 국내 인공지능 시장 동향



※ 출처: 선도형 SW R&D 추진계획, 미래창조과학부, 2014

그러나 국내의 인공지능 기술수준은 선진국 대비 약 2.4년(2015년도 ICT수준 조사보고서, IITP, 2016.2) 내지 4.4년(AI 기술과 산업의 가능성, ETRI, 2015.10) 정도의 기술 격차가 있는 것으로 파악되고 있다. 2015년도를 기준으로 인공지능 분야 국가별 누적 특허출원 건수 또한 미국 뿐 아니라 중국이나 일본에도 뒤쳐져 있다.

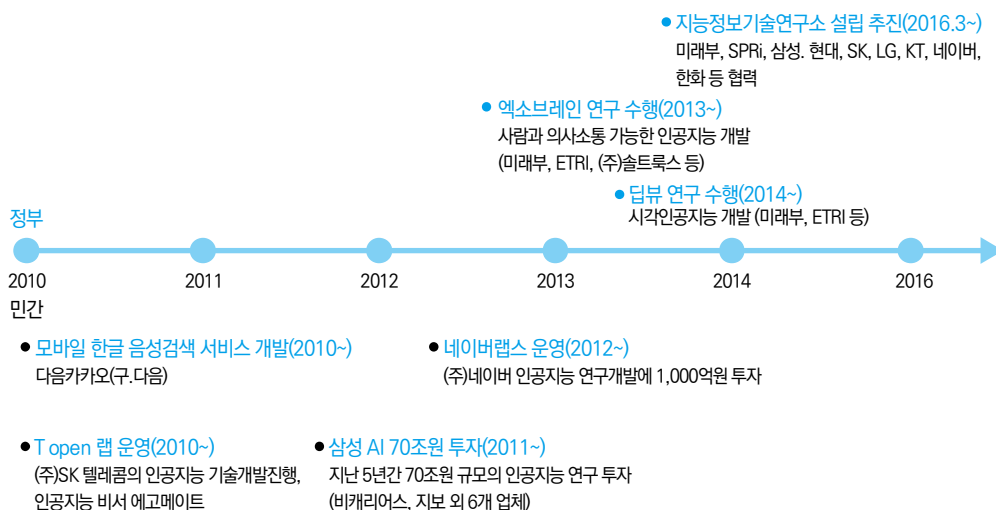
표 3-4-4 인공지능 분야 국가별 특허출원 건수

국가	미국	중국	일본	한국
건수	9,867	6,900	4,208	2,638

※출처: 팻스냅, 2015

국내에서도 지난 2010년부터 인공지능 기술의 글로벌 경쟁력 제고를 위한 노력이 일부 대기업과 정부 주도로 진행되고 있다[그림 3-4-2 참조]. 그러나 언어인지나 시각인지 등과 같이 단기간에 가시화할 수 있고 수요가 많은 분야 위주로 연구가 수행되고 있다.

그림 3-4-2 국내 인공지능 행보 대표사례



※ 출처: 한국의 인공지능 산업, SW정책연구소, 2016

정부차원에서는 2013년부터 미래창조과학부를 중심으로 인공지능기술 R&D가 본격적으로 추진되고 있다. 2015년에는 미래부의 SW R&D 전체예산 2,017억원 중 380억원(18.8%)이 인공지능 관련 분야에 투입된 바 있다. 또한 미래부의 R&D 역시 언어지능(엑소브레인 프로젝트), 시각지

능(답부 프로젝트), 기계학습(SW연구센터) 등을 집중 지원하고 있다.

표 3-4-5 미래부 인공지능 R&D 추진 현황

과제명	내용	주관기관	기간	총예산
엑소브레인 (Exobrain)	사람과 지식소통이 가능한 자연어처리 인공지능 SW 개발	KAIST, ETRI, ㈜솔트룩스 등 27개 기관	2013~2023 (4+3+3)	1,070억원
답부	사람처럼 시각을 통해 상황을 이해하는 시각 인공지능 SW 개발	ETRI, SK, 광주 과학기술원 등 29개 기관	2014~2025 (4+3+3)	약 800억원
SW 연구센터	기계학습과 관련된 방법론 및 알고리즘 연구(응용분야: 컴퓨터 비전, 데이터마ining, 뇌-컴퓨터 인터페이스)	포항공대	2012~2018	70억원
SW 스타랩	일상생활학습 기반의 인지에이전트 소프트웨어 개발	서울대학교	2015년부터 3~7년 내외	60억원
	이동 환경에서 뇌-컴퓨터 인터페이스를 위한 지능형 패턴 인식 소프트웨어 개발	고려대학교	2015년부터 3~7년 내외	60억원

※ 출처: 인공지능 실태조사, IITP, 2015

한편, 2016년 3월에 발표된 미래부의 ‘지능정보산업 발전전략’에 의하면 향후 5년간 지능정보 기술연구소 설립 등을 포함하여 지능정보산업 분야에 1조원을 투자하고, 민간에서도 2조 5,000억원 이상의 추가 투자가 발생할 수 있도록 유도할 계획이다.

2015년 현재 국내 인공지능 관련 기업은 약 60여 개로 추정되고 있다. 삼성전자, 네이버, 다음 카카오 등의 IT 업체에서는 글로벌 AI 기업에 투자하거나 자체기술 확보를 위한 노력을 시작하였고, 다수의 IT 중소·벤처기업에서도 기존의 제품에 인공지능 기술을 접목하여 자연어처리 서비스, 헬스케어, 금융, 게임 등의 분야에서 산업혁신을 위한 노력을 기울이고 있다.

표 3-4-6 국내 인공지능 기업 현황

구 분	주요기업
자연어처리 서비스	솔트룩스(아담), 다이퀘스트(마리너4), 코난테크놀로지 등
헬스케어	루닛, 뷰노, 네오파트, 디오텍, 스탠다임, 제노플렌 등
금융	위버플(스넥), 솔리드웨어, 씽크폴 등
게임	엔씨소프트(블레이드앤소울), 넷마블(콜럼버스), 넥슨(서든어택2) 등
대기업	삼성전자, 네이버, 다음카카오 등
기타	스캐터랩(감성분석), 코노랩스(개인비서), 유비파이(드론) 등

4. 인공지능 기술개발 동향

서두에서 언급한 바와 같이 인공지능은 이제 대부분의 산업부문에서 핵심요소 기술로 인식되고 있다. 이러한 현상은 2010년경 딥러닝으로 대표되는 기계학습 기술의 획기적인 발전에 기인하고 있으며, 이는 그동안 꾸준히 축적되어 온 빅데이터와 GPU(Graphics Processing Unit) 등과 같은 자료처리장치의 발전이 없었다면 불가능한 것이었다. 기계학습 기술의 발전은 인간의 대표적인 지능인 시각지능과 언어지능, 그리고 지식획득 지능의 획기적인 개선을 가능하게 하였고, 이를 통해 영상인식, 음성인식, 대용량 자료분석 등이 인간의 수준에 필적하거나 인간의 수준을 능가하게 되었다.

미국 내 글로벌 IT기업들이 주도하고 있는 인공지능 기술 개발의 주요 특징 중 하나는 개발된 기술을 경쟁적으로 공개하고 있다는 점이다. 2015년 11월에 구글은 리눅스 기반의 딥러닝 라이브러리 텐서플로를 공개하였고, 마이크로소프트도 비슷한 시점에 윈도 환경에서의 분산 기계학습 툴킷 DMTK(Distributed Machine Learning Toolkit)와 다계층 신경망 학습 툴킷 CNTK(Computation Network Toolkit)를 공개하였다. 페이스북 또한 2015년 1월에 자사의 딥러닝 라이브러리인 토치(Torch)를 공개한 바 있고, 야후는 2016년 2월에 C-언어 기반의 카페온스파크(CaffeOnSpark)를 공개하였다. 이들 미국 기업들 외에도 중국의 바이두는 2016년 1월에 음성인식 기능 위주의 라이브러리인 Warp-CTC를 공개하였고, 삼성전자도 2015년 11월에 자사의 분산 딥러닝 플랫폼 벨레스(Veles)를 공개하였다. 이와 같이 최근에 공개된 오픈소스 라이브러리 외에도 지난 10여 년간 40여 가지의 기계학습 오픈소스가 공개된 바 있다.

인공지능 기술을 선도하고 있는 글로벌 기업들이 개발된 기술을 경쟁적으로 공개하는 이유는 기술주도권을 선점하여 시장 영향력을 확대하고, 다양한 응용을 발굴하며, 사용자 확장을 통해 우수한 인재를 확보하고, 외부 연구자들의 도움으로 기술개선 효과를 도모하는 등의 장점이 있기 때문이다. 아울러 개발된 기술의 공유를 통해 누구나 쉽고 빠르게 인공지능 기술을 접할 수 있도록 함으로써 개발과정의 시행착오를 최소화하고, 집단지성에 의한 협력개발로 인공지능 분야 기술발전에 공헌할 뿐 아니라, 인공지능 기술의 잘못된 사용을 사전에 방지할 수 있는 등 공익적 측면의 장점도 있다.

5. 맺음말

2016년 3월에 있었던 알파고와 이세돌의 바둑대결 이벤트는 우리나라 국민들을 포함하여 정부, 기업, 연구소, 대학 등에 인공지능 및 기계학습 기술에 대한 폭발적인 관심을 이끌어냈다. 그러나 국내의 인공지능기술은 글로벌 수준에 비해 미약하고, 전문인력 또한 턱없이 부족한 것이 현실이

다. 그동안 정부와 민간의 투자가 거의 없었기 때문이다.

이제부터라도 산·학·연·관이 합심하여 지능정보기술의 기반을 다지고, 이를 바탕으로 우리나라 미래 신산업의 혁신을 추구해야 한다. 다행히 정부에서는 지능정보기술연구소 설립 등을 포함하여 지능정보산업 분야에 향후 5년간 1조원을 투자한다고 하였고, 미래부, 교육부, 산자부 등의 정부 부처에서는 다수의 인공지능 관련 R&D 사업 및 인력양성 사업을 준비하고 있다. 민간 기업들에서도 인공지능 관련기술 및 인력 확보를 위한 투자를 서두르고 있고, 기술력을 바탕으로 하는 다수의 스타트업도 등장할 것으로 예상된다.

지금까지 글로벌 기업들을 중심으로 발전된 인공지능 및 기계학습 기술은 대부분 공개되어 있다. 상대적으로 뒤쳐진 국내 연구자들에게는 유리한 여건임이 분명하다. 이들 공개 소프트웨어를 기반으로 독자 기술을 개발하고, 산업 전반에 지능정보기술을 접목하는 움직임이 여기저기서 많이 시도되어야 할 것이다.

제2절 빅데이터 기술

1. 개요

대용량 데이터 처리를 분산 병렬 컴퓨팅에서 처리하기 위한 목적으로 제작·발표된 맵리듀스는 현업의 많은 개발자 및 관리자들이 빅데이터 처리·분석을 위해 사용하는 프로그래밍 모델이다. 대표적인 맵리듀스 애플리케이션으로 아파치 소프트웨어 재단에서 개발 지원하고 있는 하둡이 있다.

하지만 최근에는 빅데이터 처리·분석을 위해 기존 하둡 맵리듀스(Hadoop MapReduce) 방식에서 스파크 방식으로 전환을 시도하는 기업들이 점차 증가하고 있다. 아파치 스파크 2.0이 출시되면서 빅데이터 처리·분석 분야에서도 변화의 움직임이 조금씩 나타나기 시작하였다. 본 절에서는 대용량 데이터 처리를 위한 하둡 맵리듀스와 최근 부각되고 있는 아파치 스파크를 중심으로 살펴 보도록 한다.

2. 빅데이터 처리·기술 현황

가. 하둡 맵리듀스 vs. 아파치 스파크

하둡 맵리듀스는 현재까지 가장 각광받고 있는 빅데이터 분석 및 처리를 위한 프레임워크로서 하둡에 기본 탑재되어 지금까지 많은 개발자들이 대용량 데이터를 다루는 데 사용해왔다.

하둡 맵리듀스는 1) 데이터가 있는 서버로 코드 전송, 2) 키-밸류(Key-Value) 형태의 데이터 셋 변환 처리 3) 뛰어난 병렬 처리 4) 오프라인 배치 처리 최적화 등의 장점으로 인해 현업에서 많이 활용되고 있다.

하지만 하둡 맵리듀스에는 다음과 같은 몇가지 문제점이 상존하고 있다. 1) 분산 처리를 위해서 크기는 작지만 다수의 자바 코드를 작성해야 하는 문제, 2) 독특한 동작 방식으로 인한 맵리듀스 코드의 유지보수 문제, 3) 맵과 리듀스 중간에 발생하는 대량의 I/O 문제, 4) 데이터 스키마를 표현할 방법이 마땅하지 않은 문제, 5) 어려운 문제를 해결하고자 할 때 필요한 다양한 도구 부재, 6) 성능 최적화 문제, 7) 여러 유형의 데이터를 결합하고자 할 때 방법의 문제 등이 그것이다.

맵리듀스의 문제점에도 불구하고 지금까지 산업현장에서 시간은 오래 걸리지만 서버 성능에 구애받지 않고 잘 구동되며, 특히 ETL과 같은 데이터 정제작업에 있어서는 모든 데이터를 읽어야 하므로 맵리듀스는 최적의 처리 방식으로 인정받아 왔다. 하지만 서버의 사양이 높아지고 도입비용은 낮아지면서 이러한 문제를 극복하기 위해 현업에서 논의되고 있는 것이 바로 ‘아파치 스파크’이다.

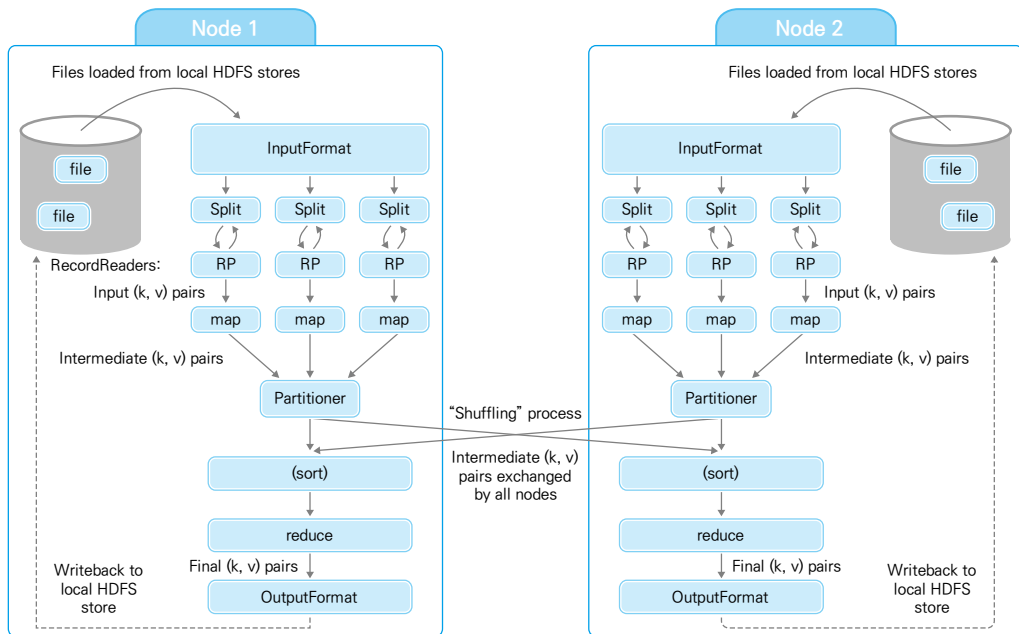
스파크 방식의 장점은 1) 맵리듀스 대비 단순한 프로그램 작성 방식, 2) 기계학습 및 복잡한 처리를 위한 다양한 API 제공, 3) 정형화된 데이터 모델의 지원, 4) 배치 분석 및 실시간 분석을

동시에 수행할 수 있는 기능 제공, 5) RDBMS, 파일, HDFS 등 다양한 데이터소스를 동시에 사용할 수 있는 능력, 6) 맵, 리듀스 중간에서 발생하는 대량의 I/O를 인메모리로 처리하여 성능 개선, 7) 하둡 없이도 동작할 수 있는 아키텍처, 8) R과 강력한 연동을 통한 분산 알고리즘 실행 지원 등을 들 수 있다.

하둡의 맵리듀스는 맵과 리듀스로 분리되어 있으며 복잡한 데이터를 처리하고자 할 때 개발자는 분산 처리를 고려하여 자바로 다수의 맵과 리듀스를 작성해야 한다. 하나의 간단한 데이터 처리가 다수의 자바 코드로 분리되어 분산 처리가 되다 보니 개발도 어렵고 유지보수도 어려운 문제가 발생한다. 스파크는 맵리듀스의 수많은 코드를 하나의 클래스로 정리할 수 있다. 따라서 개발자는 분산 처리를 고려하지 않고 아주 간단하게 코드를 작성할 수 있다. 복잡한 과정을 하나의 클래스에 모두 넣어서 처리하므로 매우 단순하다.

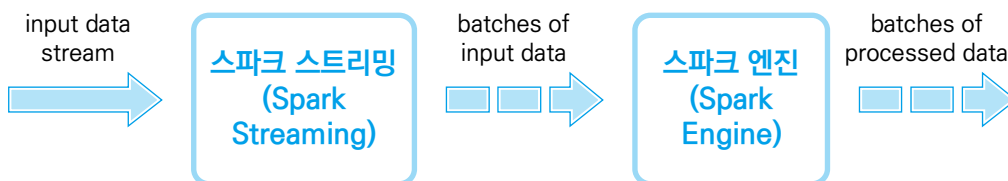
또한 맵리듀스는 [그림 3-4-3]에서 보는 것처럼 데이터를 읽고 저장하기까지 일련의 처리 방식이 표준화되어 있는 매우 단순한 프레임워크이다. 따라서 데이터 정렬, 그룹핑, 피봇팅 등 일련의 데이터를 처리하기 위해서 모든 과정을 직접 개발자가 구현해야 한다. 반면 스파크는 이미 제공되는 수많은 API를 통해서 대부분의 데이터 처리 작업을 간단히 수행할 수 있다.

그림 3-4-3 맵리듀스의 데이터 처리 흐름



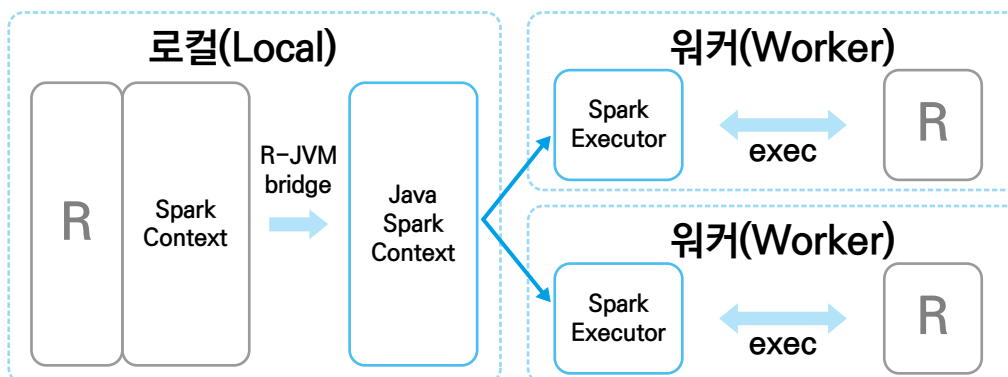
스파크는 배치 분석을 위해서 분석하는 용도로 개발되었지만 추후 실시간 처리를 지원할 수 있도록 스파크 스트리밍(Spark Streaming) 프로젝트가 진행되어 현재는 아파치 스파크에 서브 프로젝트로 같이 유지보수가 되고 있다. 스파크의 기존 개발 방법을 그대로 실시간까지 적용할 수 있다는 것이다. 스파크 스트리밍은 기존의 배치 작업을 수행하는 데 소요되는 시간을 고정된 시간 씩 끊어서 실행할 수 있고 이 시간을 아주 짧은 시간으로 정하면 마치 실시간으로 처리하는 것처럼 동작한다.

그림 3-4-4 스파크 스트리밍의 짧은 시간 동안의 배치 처리를 통한 실시간 처리



맵리듀스는 본래 하둡의 일부이므로 반드시 하둡을 구성해야만 사용할 수 있다. 이런 이유로 맵리듀스는 기본적으로 갖추어야 할 서버와 스펙이 있다. 마스터(Master) 2대, 워커(Worker)는 최소 1대 이상을 구성해야 하므로 보통 5~6대가 최소 구성이다. 반면 스파크는 하둡이 꼭 필요하지 않으므로 상황에 따라 적은 수의 서버를 사용해도 무방하다. 스파크는 최근 스파크R(SparkR)이라는 서브 프로젝트를 통해 R과 함께 사용할 수 있도록 지원하고 있다. 다음의 도식은 R에서 스파크를 호출하여 R을 다수의 워커(Worker) 노드(서버)에서 분산 처리를 수행하는 도식이다.

그림 3-4-5 R과 스파크의 통합 분석 아키텍처





3. 맺음말

지금까지 하둡 맵리듀스와 아파치 스파크에 대한 비교 분석을 통해 최근 빅데이터 처리·분석 기술 동향에 대해 살펴보았다. 특히 최근 부각되는 아파치 스파크의 기술적 특성에 대해 비중을 두어 설명해 보았다. 하둡 맵리듀스가 여전히 시장의 절대적 우위를 점하고 있으나, 아파치 스파크 방식이 나름의 강점을 바탕으로 시장에서 활용도가 점점 높아지고 있는 추세이다. 다양한 빅데이터 분석 기술의 발달로 더 빠르고, 더 편리하고, 더 다양하게 빅데이터를 처리·분석할 수 있는 환경이 갖추어지고 있다.

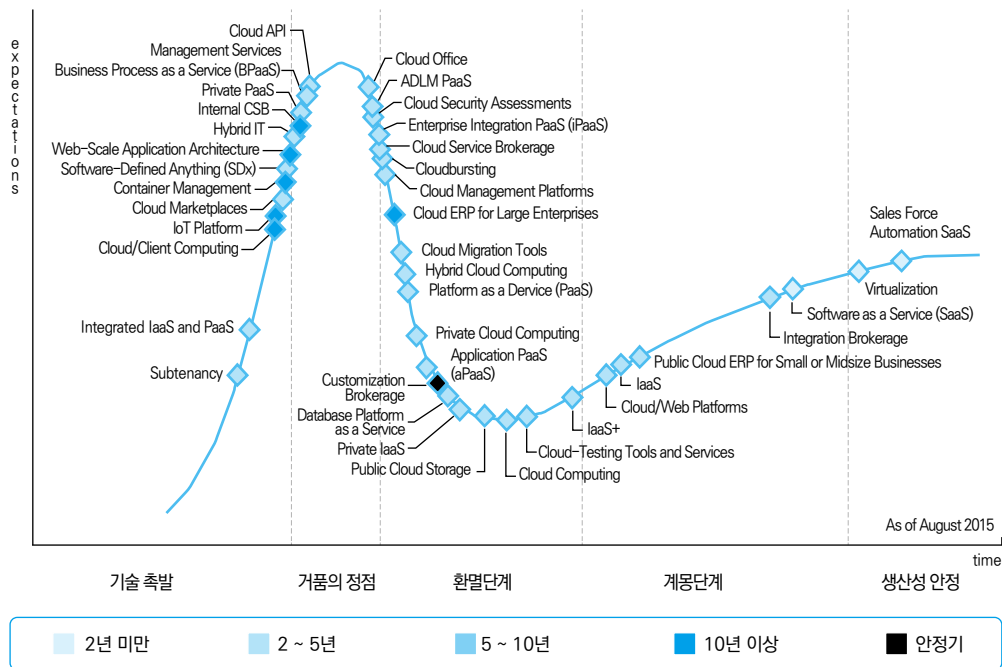
제3절 클라우드 기술

1. 클라우드 기술 동향

SaaS(Software as a Service), PaaS, IaaS(Infrastructure as a Service)는 각각 애플리케이션, 미들웨어 및 인프라를 클라우드에서 인터넷을 통해 가입형으로 제공하는 서비스이다. 클라우드 사용자 관점에서 클라우드를 기업 내에 구축하여 활용하면 프라이빗 클라우드, 글로벌 시장 대상으로 제공되는 상용 서비스를 가입하여 사용료를 지불하며 활용하면 퍼블릭 클라우드라고 한다.

아래 가트너의 하이프 사이클(Hype Cycle)에서 보듯이 Public SaaS와 Public IaaS는 이미 시장의 성숙단계에 진입하였다. Public PaaS는 아직 기술이나 제품들이 성숙단계에는 이르지 못하였으나, 벤더들간의 경쟁이 심화되면서 빠른 속도로 발전하고 있다.

그림 3-4-6 신기술 하이프 사이클



※ 출처: Hype Cycle for Cloud Computing, 가트너, 2015

프라이빗 PaaS는 아직 널리 확산되기에는 기술이나 제품의 성숙도가 낮아 IT 성숙도가 전세계적으로 가장 높은 소수의 기업들에서만 성공 사례를 찾아볼 수 있다. 프라이빗 IaaS는 프라이빗 PaaS보다는 제품의 성숙도가 다소 높아 여러 기업에서 시도되고 있으나 아직 성공한 기업보다는

실패한 기업이 더 많은 상태이다. 프라이빗 SaaS는 프라이빗 PaaS 및 IaaS에 의존하므로, 이 또한 아직은 성공사례가 많지 않다.

2. Database Platform as a Service

데이터를 퍼블릭 클라우드에 저장하여 놓고 사용할 수 있도록 하는 서비스는 PaaS 서비스에 속한다. [그림 3-4-6]에서 보듯이, DBPaaS는 퍼블릭 PaaS의 여러 종류들 중에 기술이나 서비스의 성숙도가 높은 편이다.

DBPaaS는 기업이 DBMS 라이선스를 구매하지 않고, 클라우드에서 제공하는 DBMS를 인터넷 서비스 형태로 활용할 수 있도록 한다. DBPaaS 벤더는 DB 운영 및 관리의 일체 서비스(컨피규레이션, 성능 모니터링, 튜닝, 가용성, 백업 등)를 제공한다. 따라서 DBPaaS 사용 기업은 DBMS 관리 및 최적화에 신경 쓸 필요 없이, 애플리케이션에서 DB를 연결하여 쿼리 및 업데이트 트랜잭션을 수행하기만 하면 된다.

클라우드 서비스의 필수 속성들을 모두 갖추어야만 진정한 DBPaaS로 간주된다. 즉 많은 기업이 한 개의 DBMS 인스턴스를 공통으로 사용하되, 각 사용 기업의 DBMS 컨피규레이션, 저장된 데이터, 트랜잭션 등을 상호 독립된 상태로 개별 지원할 수 있는 멀티테넌트 서비스이어야 한다. 또한 사용 기업들의 데이터베이스 워크로드가 증감함에 따라 필요한 컴퓨팅 자원을 신축성 있게 자동으로 할당하고, 컴퓨팅 용량을 필요에 따라 자동으로 무제한 증가시켜 나갈 수 있어야 한다. 이를 클라우드의 탄력성 및 확장성이라 한다. 또한 사용 기업들의 컴퓨팅 자원 사용량(또는 사용자수)을 자동 측정하여 사용료를 자동 과금하는 기능도 갖추어야 한다. 가장 중요한 것은 셀프 서비스 속성이다. 즉 사용 기업이 클라우드 벤더의 도움 없이 셀프 서비스로 가입하고, DBMS를 설정하고, 필요 자원을 즉시 프로비저닝 받아 활용할 수 있어야 한다.

DBPaaS 서비스를 제공하는 벤더들도 늘어나고 있다. DBPaaS 서비스 중에는 ACID(Atomicity, Consistency, Isolation, Durability) 속성을 보장하는 R DBMS 서비스도 있고, 최근 확산되고 있는 NoSQL DBMS 서비스도 있다.

DBPaaS를 제공하는 벤더들로는 아마존, 클러스트릭스(Clustrix), 엔터프라이즈DB (EnterpriseDB), 구글, 헤로우(Heroku), IBM, 마크로직(MarkLogic), 마이크로소프트, 엠랩(mLab), 누오DB(NuoDB), 오브젝트로켓(ObjectRocket), 오라클, 오케스트레이트(Orchestrate), 오리엔트DB(OrientDB), 랙스페이스(Rackspace), 레디스 랩스(Redis Labs), SAP, 스노우플레이크(Snowflake), 테라데이터, 트랜스라티스(TransLattice) 등이 있다. 이들은 R DB뿐 아니라 DW와 각종 NoSQL DB를 제공하고 있다.

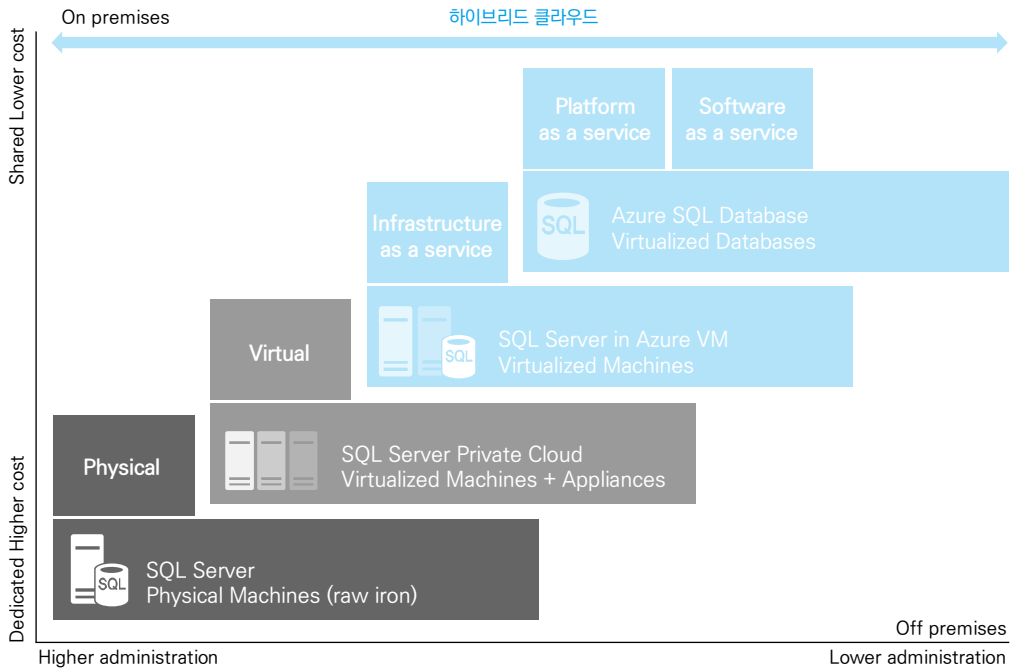
DBPaaS와 유사해 보이나 사실상 상이한 서비스가 있다. 이는 기업이 자체 보유한 DBMS를 비추얼머신 이미지화하여 퍼블릭 IaaS에 호스팅시키는 경우이다. DBPaaS와 DBIaaS 두 경우를

함쳐 보통 DBaaS 또는 클라우드 DB라고 한다.

거의 모든 관계형 및 NoSQL DBMS 벤더들이 DBIaaS를 지원하고 있어 DBMS를 쉽게 아마존 EC2, 마이크로소프트 애저 등 IaaS에 설치 운영할 수 있다. DBIaaS 서비스를 제공하는 IaaS 벤더들은 DBPaaS 벤더와는 달리 DBMS 자체의 운영 관리 서비스는 제공하지 않으며, 하드웨어, 네트워킹 등 인프라 관련 서비스만 제공한다.

기업들은 DBMS들을 자체 데이터센터뿐 아니라 퍼블릭 IaaS에도 설치하고 DBPaaS도 활용하는 소위 하이브리드 클라우드 전략을 추구할 수 있다. IBM 클라우드런트, 마크로직, 마이크로소프트 애저 등이 하이브리드 솔루션을 제공하고 있다. [그림 3-4-7]은 마이크로소프트 SQL 서버 DBMS를 자체 데이터센터에 설치, 프라이빗 IaaS에 설치, DBIaaS에 설치, 또는 DBPaaS로 활용할 수 있음을 나타내고, 각 경우의 DB TCO 및 관리 부담 정도를 비교하여 보여준다.

그림 3-4-7 DBMS의 하이브리드 클라우드 전략



※ 출처: Choose a Cloud SQL Server Option, 마이크로소프트, 2016

3. Data as a Service(DaaS)

Data as a Service(DaaS)라는 용어도 있는데, 데이터 서비스(Data Service)라는 용어와 혼용하여 쓰인다. DaaS는 클라우드와 직접 연관된 개념은 아니다. DaaS는 많은 사용조직이 공통으로 필요로 하는 공용 데이터를 인터넷 또는 인트라넷을 통해 서비스 형태로 제공하는 것을 의미한다. REST 서비스와 같은 인터넷 기반의 국제표준 데이터 서비스 프로토콜을 이용하여 기업들에게 필요한 공용 데이터를 제공하는 상용 서비스들이 있다. 한편 기업 내에서 애플리케이션들을 서비스 지향 아키텍처(SOA, Service-Oriented Architecture)로 구축하면서, 부서 간 공통으로 사용하는 데이터(즉 Master Data)를 중앙저장소(Central Repository)에서 DaaS 개념의 공통서비스로 공급하는 것도 가능하다.

마스터데이터 관리(MDM, Master Data Management), 고객 데이터 통합(CDI, Customer Data Integration), PIM(Product Information Management) 등에서 DaaS 개념 및 기술을 구현하고 있다. 기업이 고객, 제품, 자산 등과 같은 전사 공용 데이터를 마스터데이터 허브에 저장하여 두고, 어느 애플리케이션에서든 “Get Customer”와 같은 간단한 서비스 호출을 통해 필요한 데이터를 가져갈 수 있도록 한다.

인텔은 오라클의 DaaS 솔루션을 이용함으로써, 고객 관련 방대한 외부 데이터를 활용할 수 있게 되었다. 이를 고객 세분화 모형(일종의 Predictive Analytics 모델)을 개선하는데 활용하였으며, 그 결과 마케팅 캠페인의 원가를 75% 감축할 수 있었다.

DaaS 기술과 아키텍처는 기업의 정보 표준화, 통합 및 공유, 데이터 품질관리 등을 위해 중요한 기술이며, 기업이 정보시스템들을 SOA 아키텍처로 구현해 가는 데 중요한 기초를 제공한다. DaaS를 제공하는 중앙데이터저장소(Central Data Repository)를 퍼블릭 또는 프라이빗 클라우드에 구축하는 것이 충분한 이점이 있을 것으로 보인다. 그러나 DaaS의 클라우드 서비스화가 아직 많은 기업이 보편적으로 적용할 수 있도록 성숙된 기술은 아니다.

4. 클라우드 DB 도입을 위한 의사결정

DBMS 사용 기업 입장에서 종전처럼 DBMS 제품 라이선스를 구매할 것인가, 아니면 구매하지 않고 DBPaaS 서비스에 가입하여 사용료만 지불할 것인가? DBMS 제품 라이선스를 구매한 경우에도 DBMS를 자체 데이터센터 또는 전통적인 인프라 호스팅 벤더의 데이터센터에 설치할 것인가, 아니면 DBIaaS 벤더의 클라우드센터에 설치할 것인가? 이러한 의사결정을 할 때 어떤 요소들을 고려해야 하나?

자체 정보시스템 구축 운영이 어려운 중소기업이나 창업기업들은 애플리케이션과 DB를 통합 제공하는 SaaS를 선호할 것이다. 기업이 정보시스템과 그에 필요한 DB를 자체 개발 활용하는 경우 [표 3-4-7] DBMS 및 인프라를 확보하는데 여러 방안이 존재하게 되었다. 최적의 방안을 선별할 때 먼저 방안별로 총소유비용(TCO, Total Cost of Ownership)을 측정하여 비용 측면의 우열을 판단해 봐야 한다. DBMS 또는 데이터센터를 보유한 기업은 구매 내지 구축 비용이 투자되었기 때문에 클라우드 도입이 비용 면에서 불리할 수 있다.

클라우드 DB 사용 시, 애플리케이션과 DB간 양방향 데이터 트랜스퍼의 양이 많을 경우 과다한 통신비용이 사용료에 반영되어 과금될 수 있음도 간과하면 안 된다. 특히 DW 경우 추출·변환·적재 작업에 많은 양의 데이터 전송이 필요할 수 있으므로 클라우드 DB 사용 시 통신비와 트랜잭션 반응속도에 대해 사전 검토하는 것이 중요하다.

표 3-4-7 DBMS 및 인프라 확보 방안

DBMS 확보 방안	인프라 확보 방안	총소유비용 (TCO)
DBMS 자체보유	데이터센터 자체보유	- 인프라 구축, 운영, 관리 비용 - DBMS 구매, 유지보수, 운영, 관리 비용
	인프라 호스팅 서비스 활용	- 인프라 호스팅 서비스 비용 - DBMS 구매, 유지보수, 운영, 관리 비용
	DBaaS 활용	- IaaS 사용료 (통신비용 포함) - DBMS 구매, 유지보수, 운영, 관리 비용
DBPaaS 활용		DBPaaS 사용료 (통신비용 포함)

클라우드 DB 도입 시, 비용 못지 않게 혜택 측면이 더욱 중요할 수 있다. NoSQL, 인메모리 DB 등 신기술 적용을 통한 경영혁신을 도모하는 경우 DBPaaS 서비스를 활용하여 출시 소요시간을 줄이고, 내부 기술역량의 부족을 극복할 수 있다. 기존의 사일로(Silo) DB 인스턴스들을 전사 표준 메타데이터 기반으로 통합하여 여러 조직단위 간에 공유하고자 하는 경우 DBPaaS로 전환을 고려할 수 있다. 기업 내 DB 운영 및 관리 역량 미흡으로 가용성, 성능, 신뢰성이 취약하여 정상적인 경영활동을 저해하는 경우 DBPaaS 활용을 통해 개선을 도모할 수 있다. 비즈니스 지속성 및 재난복구를 위한 DB 백업 및 복구 능력을 제고해야 할 경우에도 DBPaaS 도입을 검토할 수 있다.

비용/혜택 분석 외에 클라우드 사용과 관련된 일반적인 고려사항들을 모두 검토해야 한다. 법규 제상의 데이터 저장 장소 제한, 정보보안의 중요도, DB 워크로드의 규모 및 변동성, 트랜잭션 반응속도에 대한 요구조건, 가용성 요구조건, DB 운영에 관한 내부통제의 중요성 등을 주요 고려사항으로 들 수 있다.

지금까지 DBPaaS를 활발히 활용하는 사례들을 보면 DB에서 네트워크를 통해 전송해야 할 데

이터 양이 많지 않고, 정보보안 문제가 심각하지 않은 경우들이다. 특히 애플리케이션 개발 프로젝트에서 DBPaaS를 활용하는 경우가 많다. DB 사용에 있어 특수한 컨피규레이션, 특수한 보안 통제, 특수한 DB 관리가 필요한 경우에는 DBPaaS보다는 DBIaaS가 적합할 수 있다. 일반적인 상황에서는 DBPaaS를 사용하면 패칭(Patching), DB 컨피규레이션, 튜닝 등 데이터베이스의 관리 업무의 품질을 제고할 수 있다. DBIaaS를 사용하는 경우에는 백업, 가용성 보장 등 전문 기술을 요하는 인프라 서비스 플래닝 업무와 인프라 자원에 대한 투자를 전문 벤더에 위임하게 되어 DB 인프라 관리의 비용, 효율 및 품질을 향상시킬 수 있다.

퍼블릭 DBIaaS나 DBPaaS를 사용하는 경우 모두, 사용자 및 애플리케이션 서버의 지리적 위치와 DB서버가 위치하는 클라우드센터와의 거리에 따라 트랜잭션 반응속도가 크게 영향을 받는다. 따라서 벤더 선정 이전에 파일럿 프로젝트를 통해 성능 테스트를 하는 게 중요하다. 가용성 관리 부문에서도 SLA상 가용성 지원 옵션들, 백업 사이트의 자동 선정, 복구 절차 및 소요시간 등을 세밀히 검토해야 한다.

DBPaaS가 아직은 애플리케이션 개발 및 테스트를 위한 DB의 일시적 사용에 주로 활용되고 있으나, DBPaaS 기술 및 서비스 성숙도가 올라감에 따라 점차 기업의 기간 정보시스템에 활용되는 사례가 증가하고 있다. 특히 애플리케이션의 개발 시 DBPaaS가 선호되는 주요 이유는 DB 셋업 시간이 크게 단축되기 때문이다. 가트너는 DBPaaS 시장이 2013~2018년 기간 중 연 39.7%로 빠른 성장을 보일 것으로 예측하고 있다.

제4절 사물인터넷 기술

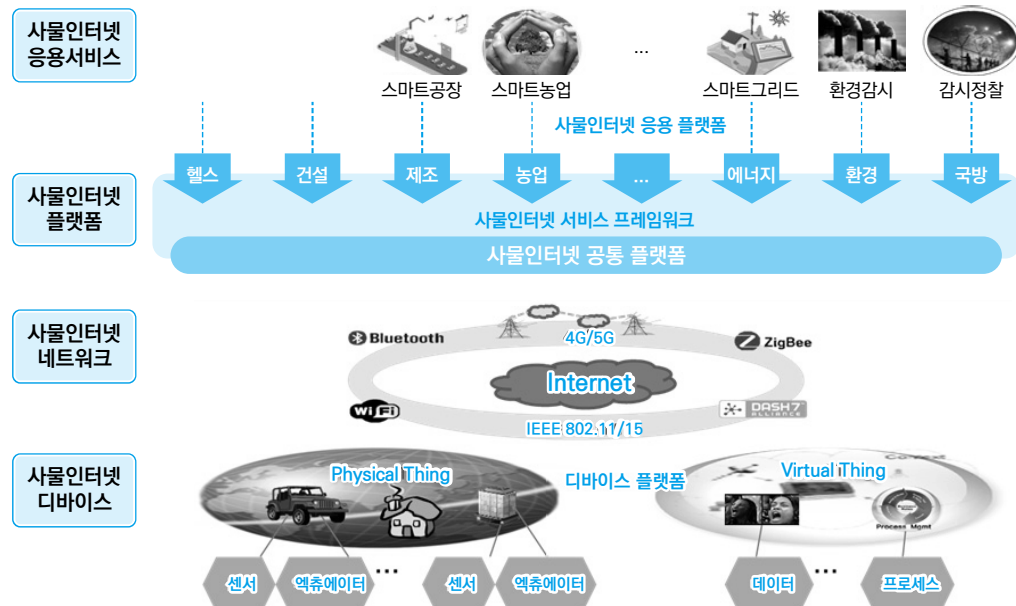
1. 개요

가. 사물인터넷 기술 정의

사물인터넷 기술은 인터넷을 기반으로 다양한 물리적 및 가상의 사물들을 연결하여 언제 어디서나 상황에 맞는 최적의 서비스를 제공하기 위한 글로벌 서비스 인프라로 정의되며, 사람·사물·공간 등 모든 것들(things)이 인터넷으로 서로 연결되어, 모든 것들에 대한 정보가 생성·수집되고, 공유·활용되는 기술을 말한다. 2005년 국제전기통신연합(ITU, International Telecommunication Union)이 처음으로 사물인터넷의 개념을 정립하였으며, 사물인터넷 서비스·사물인터넷 플랫폼·사물인터넷 네트워크·사물인터넷 디바이스·사물인터넷 보안 등의 구조로 정의된다.

사물인터넷 서비스는 물리적 및 가상의 사물과 연계, 협업하여 지능형 서비스를 제공하는 사물인터넷 플랫폼, 모든 사물을 인터넷을 통해 상호 연결하여 소통하는 사물인터넷 네트워크, 사물을 지능화시켜 스마트 인터랙션을 제공하는 사물인터넷 디바이스, 프라이버시 보호와 안전한 시스템 운영을 보장하는 사물인터넷 보안 등을 통해 제공된다.

그림 3-4-8 사물인터넷 개념도



※ 출처: 사물인터넷 R&D 추진계획, 미래창조과학부, 2014

나. 기술의 범위

사물인터넷 기술의 범위는 센싱기술, 유무선 통신 및 네트워크 인프라기술, 사물인터넷 서비스 인터페이스 기술 등으로 정의될 수 있다.

1) 센싱기술

일반적인 센싱기술은 전통적인 온도·습도·열·가스·속도·조도·초음파 등과 같은 물리량을 다양한 방법을 이용해 측정하는 기술에서부터 원격 감지, 레이더, 위치, 모션, 영상 센서 등 유형 사물과 주위 환경으로부터 정보를 얻을 수 있는 물리적 센서를 포함한다. 물리적인 센서는 응용 특성을 좋게 하기 위해 표준화된 인터페이스와 정보 처리 능력을 내장한 스마트 센서로 발전하고 있으며, 이미 센싱한 데이터로부터 특정 정보를 추출하는 가상 센싱 기능도 포함되며 가상 센싱 기술은 실제 IoT 서비스 인터페이스에 구현된다. 기존의 독립적이고 개별적인 센서보다 한 차원 높은 다중(다분야) 센서기술을 사용하기 때문에 한층 더 지능적이고 고차원적인 정보를 추출할 수 있다.

2) 유무선 통신 및 네트워크 인프라 기술

사물인터넷에서 네트워킹 기술은 분산된 환경에서 존재하는 다양한 디바이스들의 물리적인 연결을 수행하는 유무선 네트워킹 기술로서, WPAN(Wireless Personal Area Network), WiFi(Wireless Fidelity), 3G/4G/LTE(Long Term Evolution), 블루투스(Bluetooth), 이더넷(Ethernet), BcN (Broadband convergence Networks), 위성통신, 마이크로웨어(Microware), 시리얼 통신, PLC (Programable Logic Control) 등 인간과 사물, 서비스를 연결시킬 수 있는 모든 유·무선 네트워크를 의미한다.

3) 사물인터넷 서비스 인터페이스 기술

사물인터넷 서비스 인터페이스 기술은 사물인터넷의 주요 구성 요소(인간·사물·서비스)를 통해 특정 기능을 수행하는 응용서비스와 연동하는 기술로서, 사물인터넷의 다양한 서비스 기능을 구현하기 위해서는 1) 정보의 센싱, 가공, 정형화, 추출, 처리 및 저장기능 등 ‘검출정보 기반 기술’ 2) 위치판단 및 위치확인 기능, 상황인식 및 인지기능 등 ‘위치정보 기반 기술’ 3) 정보보안 및 프라이버시 보호기능, 인증 및 인가 기능 등 ‘보안 기능’ 4) 온톨로지 기반의 시맨틱 기술을 통해 다양한 서비스를 제공할 수 있는 인터페이스 역할을 수행할 수 있어야 한다. 이외에도 오픈 센서 API, 가상화, 프로세스 관리, 오픈 플랫폼 기술, 미들웨어 기술, 데이터마이닝 기술, 웹 서비스 기술, 소셜 네트워크 서비스 등 서비스 제공을 위해 인터페이스(저장·처리·변환 등) 역할을 수행할 수 있어야 한다.

2. 사물인터넷 플랫폼 기술

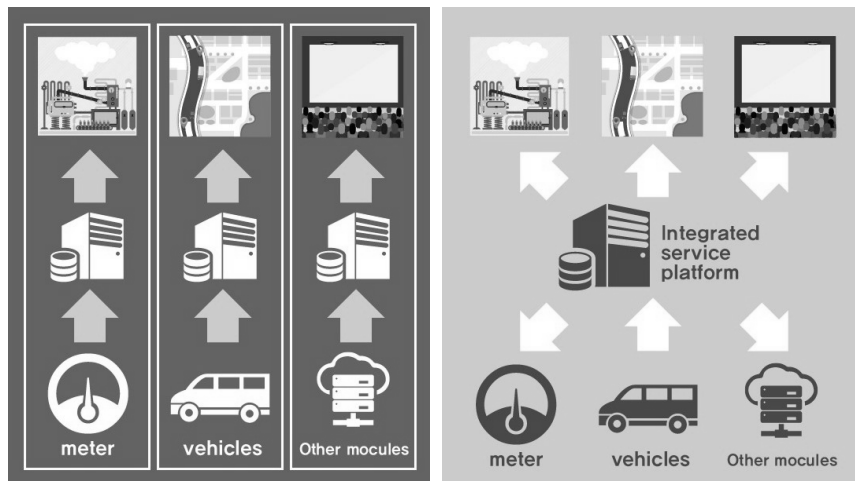
사물인터넷의 미래 서비스 시장은 대규모 사물들과 대용량 데이터를 효율적으로 처리할 수 있는 개방형 사물인터넷 플랫폼을 기반으로 크게 확산될 것으로 전망되고 있다. 이러한 사물인터넷 플랫폼의 기본적인 기능은 상호 이질적인 센서 디바이스들에 대한 연결 방법 제공, 센서 디바이스들에 대한 제어기능 제공, 대규모 센서 디바이스들로부터 수집되는 정보에 대한 효율적인 저장·관리 기능 제공, 수집된 대용량 정보를 효율적으로 검색·분석·표출하기 위한 기능 제공, 웹 등을 이용한 개방형 서비스 제공으로 볼 수 있다.

가. 기술의 정의

실세계에 존재하는 사물들과 네트워크로 상호 연결하여 사람과 사물, 사물과 사물끼리 언제 어디서나 소통하게 하여 사물들의 데이터를 수집하거나 사물에 대한 제어방법을 제공하고 궁극적으로 수집한 데이터를 중심으로 지능적인 서비스를 사람들에게 제공하기 위한 서비스 프레임워크 기술이다.

초기의 사물인터넷은 해당 서비스나 응용 도메인별로 독립적인 플랫폼 위주의 수직적 모델로 구축되었으나 시간이 흐를수록 서비스나 응용 도메인의 경계를 허무는 통합 플랫폼 위주의 수평적 모델로 접근하고 있으며 그를 위한 기능도 발전하고 있다. 주의할 부분은 수평적 모델이 하나의 플랫폼으로 통합되는 것을 의미하는 것이 아니라 표준을 전제로 하는 복수의 플랫폼들이 인터페이스의 경계가 낮아짐을 말한다.

그림 3-4-9 수직적 모델(좌)과 수평적 모델(우)의 사물인터넷 플랫폼



※ 출처: 스마트과학관, <http://terms.naver.com/entry.nhn?docId=3386835&cid=58369&categoryId=58369>

나. 기술의 특징

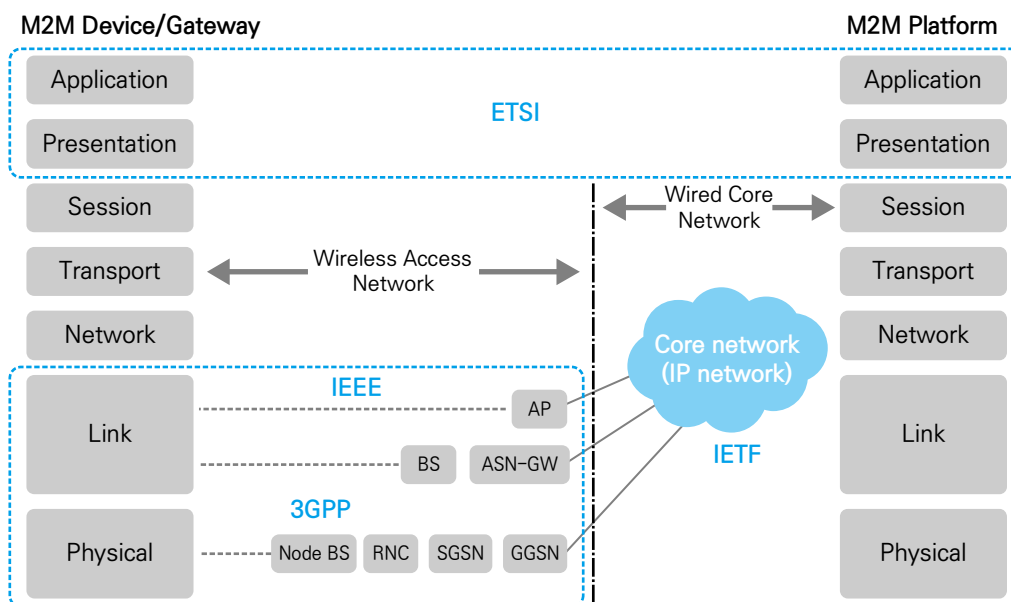
- 1) **이기종 센서 디바이스 연결 및 제어:** 이기종 센서 디바이스들을 효율적으로 연결하기 위한 사물인터넷 플랫폼의 핵심기술로서, 다양한 디바이스 플랫폼(아두이노(Arduino), 라즈베리 파이(Raspberry Pi), 비글보드(BeagleBoard), 인텔이디슨(Intel Edison) 등), 네트워크 인 프라(직비(ZigBee), 지웨이브(Z-Wave), 블루투스, 와이파이 등 통신 프로토콜(MQTT, HTTP, SOAP, TCP/IP 등)을 이용하는 센서 디바이스들에 대한 연결방법을 제공하는 것이다.
- 2) **사물정보 수집 및 저장:** 센서 디바이스가 설치된 사물들의 수가 증가할수록 사물들이 생성하는 데이터가 기하급수적으로 증가하게 되고, 데이터 형식도 텍스트·스트림·오디오·정지영상·동영상 등과 같이 다양화되고 있다. 사물인터넷 플랫폼은 이와 같이 대용량이면서 다양한 형식을 가지고 있는 센서 데이터를 효율적으로 수집하고 저장하기 위한 방법을 제공한다. 특히 실시간 처리가 필요한 데이터에 대해서는 데이터베이스 기반 데이터 저장관리 방법 등을 제공하며, 대규모 데이터 처리를 위해 클라우드 인프라 기반의 분산 빅데이터 저장관리 방법을 제공하기도 한다.
- 3) **사물 정보 검색·분석·표출:** 사물들로부터 광범위하게 수집되거나 축적된 정보를 분석하여 지능형 서비스를 제공하기 위한 방법을 제공한다. 스트림 처리를 포함한 실시간 분석 및 배치 분석 기능을 제공하며, 축적된 데이터 규모에 따라 빅데이터 분석 기능도 제공할 수 있다. 사물인터넷 서비스에 따라서 필터링, 통계, 데이터마이닝, 시맨틱 분석 등의 다양한 분석기법들이 제공될 수 있다. 최근에는 사물정보 분석 결과의 효과적인 가시화를 위해 구글맵 서비스와 같이 지도 매핑 기능을 융합한 히트맵(Heatmap) 분석 기능도 활발히 이용되고 있다.
- 4) **사물 정보의 개방형 웹 서비스:** 사물인터넷 플랫폼은 서비스 개발을 효율적으로 지원하기 위해 자신이 보유한 기능들을 오픈 API를 통해 외부에 노출할 수 있어야 한다. 사용자들은 이러한 오픈 API를 기반으로 사물인터넷 플랫폼 기능 활용 또는 외부 시스템과의 매시업 작업들을 손쉽게 이용함으로써 사물인터넷 서비스를 효율적으로 만들 수 있다.

3. 사물인터넷 기술 표준화

사물인터넷은 센서 디바이스, 네트워크, 플랫폼 등의 다양한 기술들이 통합되고, 다양한 유형의 사물정보 및 데이터들이 상호 연계되어야 비로써 사용자가 원하는 서비스가 가능해진다. 그러므로 사물인터넷에서 이러한 기술과 정보에 대한 상호운용을 보장할 수 있는 표준화는 매우 중요하다고

할 수 있다. 예를 들어, 센서 디바이스의 통신 프로토콜, 네트워크 방식, 플랫폼의 오픈 API 등에 대해 단일 표준안이 만들어지는 경우, 사물인터넷 서비스 개발이 훨씬 용이할 것이다. 그러나 현실은 ITU-T, ISO/IEC, IETF, IEEE, ETSI, OGC, oneM2M 등의 다양한 국제 표준기구들에서 이미 사물인터넷에 관련한 다양한 표준들이 각각 발표되고 있다. 이에 현 시점부터 모든 사물인터넷에 동일하게 적용할 수 있는 하나의 획일적인 표준안을 만드는 것은 어려움이 있다고 판단되며, 다양한 사물인터넷 표준들 간의 상호 연계가 가능하도록 기존 표준들을 개방형 아키텍처 기반으로 수정·확장하거나 새로운 연계 표준을 만드는 것이 현실적 대안이다.

그림 3-4-10 사물인터넷 표준기관별 표준범위



※ 출처: ETRI, 2015

현재 사물인터넷 관련 국제 표준은 공식적인 국제표준화 기구인 ITU-T, ISO/IEC JTC1과 사실(De-facto) 표준화 기구인 IETF, IEEE, ETSI, OGC, oneM2M 중심으로 다양한 표준안이 발표되고 있다. 국내의 경우 단체 표준 기구인 TTA와 사물인터넷 포럼에서 국내 사물인터넷 표준안 개발·적용과 국제 표준 활동에 대응하고 있다.

4. 사물인터넷 서비스 확산 방향

현재 사물인터넷은 다양한 분야로 서비스 창출을 예고하고 있다. 우선 개인의 상태, 취향 및 주변 환경을 감지하는 기술은 개인 맞춤형 헬스케어 및 웰니스 서비스, 개인 안전/보호 서비스, 스마트 홈, 스마트 오피스와 같이 사람을 편리하게 보조하는 다양한 서비스를 창출할 수 있을 것으로 기대된다. 기업들은 자동차, 선박, 로봇, 에너지, 건설, 제조 등과 같은 주력 산업 분야를 모니터링하고 제어할 수 있는 사물인터넷 기술을 이용하여 스마트 팩토리, 스마트 그리드, 스마트 카, 사물인터넷 기반 물류 최적화 서비스 등과 같이 산업 경쟁력을 크게 향상시킬 수 있는 서비스를 이용할 수 있을 것으로 기대된다. 정부의 경우는 재난재해 대응, 기상정보 예측, 스마트 시티, 스마트 교육, 스마트 팜, 스마트 하이웨이, 보건 등의 다양한 공공부문 서비스에 사물인터넷 기술을 적용하여 서비스 만족도를 크게 개선할 수 있을 것으로 전망된다.

그림 3-4-11 사물인터넷 서비스

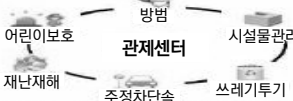
● (개인 IoT) 사용자 중심의 편리하고 쾌적한 삶

Car as a Service  차량을 인터넷으로 연결 → 안전하고 편리한 운전 ※ (예시) 긴급구난 자동전송, 무인자율주행 서비스 등	Healthcare as a Service  심장박동, 운동량 등 IoT 정보 제공 → 개인 건강 증진 ※ (예시) 심장박동 케어, 건강 팔찌 케어 서비스 등	Home as a Service  주거환경 IoT 통합 제어 → 생활 편의, 안전성 제고 ※ (예시) 가전-기기 원격제어, 홈 CCTV 서비스 등
--	---	---

● (산업 IoT) 생산성 향상 및 신 부가가치 창출

Factory as a Service  공정분석 및 시설물 모니터링 → 작업 효율 및 안전 제고 ※ (예시) 제조설비 실시간 모니터링, 위험물 감지·경보 서비스 등	Farm(&Food) as a Service  생산·가공·유통 IoT 접목 → 생산성 향상 및 안전유통체계 ※ (예시) 스마트 팜·축사·양식장, 식품 생산유통이력 정보 제공 서비스 등	Product as a Service  주변 생활제품의 IoT 접목 → 고부가 서비스 제품화 ※ (예시) 식습관 관리 포크, 심장박동음 전달 베개, 행동패턴 분석 신발 등
---	--	--

● (공공 IoT) 살기 좋고 안전한 사회 실현

Public Safety as a Service  CCTV, 노약자 GPS 등 IoT 정보제공 → 재난·재해 예방 ※ (예시) 어린이/노인 안심이, 재난재해 예보 서비스 등	Environment as a Service  대기질, 쓰레기양 등 IoT 정보제공 → 환경오염 최소화 ※ (예시) 스마트 환경정보 제공, 스마트 쓰레기통 서비스 등	Energy as a Service  에너지 관련 IoT 정보제공 → 에너지 관리 효율성 증대 ※ (예시) 스마트 건물에너지 관리, 스마트 미터, 스마트 플러그 서비스 등
--	--	--

※ 출처: 사물인터넷 R&D 추진계획, 미래창조과학부, 2014



5. 맺음말

사물인터넷 기술은 사회 전반에 걸쳐 다양한 서비스를 창출할 것으로 예상되며, 세계 최고 수준의 스마트폰과 유무선 인터넷 인프라 보급률을 보유한 우리나라는 사물인터넷 서비스 관련 선도국가 될 수 있을 것으로 기대된다. 다만, 현재 선진국과 1~3년의 기술 격차를 가지는 사물인터넷 통합 플랫폼, 빅데이터, 클라우드 인프라의 소프트웨어 기술에 대한 중요성을 인식하고, 사물인터넷 서비스 활성화를 저해하고 있는 법과 제도적 문제는 앞으로 개선해야 될 것으로 판단된다.

“빅데이터와 창조경제”



2015 우수 DB인 상 데이터컨설팅 이노베이터 수상 / ㈜투이컨설팅 **이진우** 부사장
Expert Column

최근 빅데이터에 대한 관심이 높다. IT 전문가뿐 아니라 일반인도 이제 빅데이터란 용어에 익숙하고, 일간지도 빅데이터에 대한 기사를 심심치 않게 보도하고 있다. 이와 같이 빅데이터에 대한 사회의 관심은 높지만, 빅데이터를 제대로 활용하는 기관은 많지 않다. 이는 아직 빅데이터에 대한 정확한 특성이나 활용 방안이 정착되지 않았기 때문일 것이다. 빅데이터를 마법 상자나 도깨비 방망이로 인식하고 있는 경우도 많다. 물론 빅데이터를 잘 활용하면 엄청난 부가가치를 창출하는 새로운 아이템을 발굴해 낼 수 있다. 하지만 방대한 데이터를 대상으로 어디서 무엇을 찾아야 하는지를 알아내기가 쉽지만은 않다.

사실 빅데이터는 오래전부터 우리 곁에 있었다. 산업화 사회가 성숙되어 정보화 사회로 전환되던 시점에 이미 인력으로 처리할 수 없는 방대한 데이터가 여러 시스템에 쌓여 있었던 것이다. 그래서 데이터 작업은 자동화되었고, 데이터 분석은 통계적 처리를 통하지 않고는 수행할 수 없는 시대에 살고 있다. 정보화 시대라고 하지만 우리는 역설적으로 정보의 홍수로 인한 어려움에 처해 있다. 데이터의 분량이 늘어날수록 정보 활용이 증가하리라는 기대와는 달리, 데이터에 대한 통제가 갈수록 어려워져서 고민에 빠져 있다. 이제는 데이터의 분량이 기가바이트, 테라바이트를 넘어 페타바이트나 엑사바이트라는 0의 숫자를 헤아리기도 어려운 방대한 규모로 증가하고 있다. 엑사바이트라면 0이 18개나 되는 100경이라는 엄청난 크기이다.

데이터의 급속한 증가는 데이터 활용을 제약하는 요인이다. 전에는 간단한 데이터 조합이나 처리를 통해 손쉽게 도출하던 분석 자료들이 이제는 시간도 오래 걸릴 뿐 아니라, 정확성도 떨어지고 있다. 하지만 무엇이 어떻게 잘못되었는지를 파악하기가 점점 더 어려워지고 있다. 그러므로 빅데이터로 무엇을 할 것인가도 중요하지만, 유효한 결과를 도출하려면 빅데이터를 어떻게 관리해야 하는지가 더 절박한 과제이다. 빅데이터는 신밧드의 모험에 나오는 금은보화가 넘치는 보물 창고로 인식될 수도 있지만, 사실은 고물로 가득 차 있는 어두운 동굴 속에서 숨겨진 골동품을 찾아야 하는 힘들고 고단한 작업에 더 가깝다고 보아야 한다.



빅데이터를 유용하게 활용하기 위해서는 빅데이터에 대한 체계적인 관리가 선행되어야 한다. 기하급수적으로 늘어나는 데이터를 적절히 분류해서 관리하고 필요한 시점에 쉽게 찾을 수 있는 환경이 갖추어져 있어야 한다.

대부분의 교수 연구실은 수많은 책들이 여기저기 쌓여 있다. 오래된 책에서부터 신간까지 책을 놓을 수 있는 장소에는 예외 없이 다양한 책들이 다소 무질서하게 흩어져 있다. 연구실은 좁고 소장하고 싶은 책은 많기 때문에 더 넓은 연구실로 이전하더라도 금방 책 더미 속에 빠지게 된다. 그럼에도 불구하고 연구에 필요한 책은 요령껏 찾아서 활용한다. 하지만 다른 사람이 책을 건드리려 하면 질책을 한다. 위치가 변경되면 도저히 책을 찾을 수 없기 때문이다. 만일 많은 교수들이 공유하는 도서관 크기 만한 공동의 연구실을 만든다면 어떤 일이 발생할까? 처음에는 넓은 공간을 활용할 수 있어서 모두들 행복해 할 것이다. 그러나 시간이 지나면 끔찍한 일이 발생하게 된다. 너도나도 여기저기에 책을 갖다 놓고, 사용하는 교수가 늘어나면서 공간이 부족해지고, 남의 책도 뒤적거리면서 섞이게 되면, 아무도 자기가 원하는 책을 찾을 수 없는 혼란 상태에 빠지게 된다. 문제 해결을 위하여 다시 별도의 공간을 확보하여 두 번째 공동 연구실을 만들어도 문제를 가속시킬 뿐이지 해결책이 되지 못한다. 결과적으로는 무수히 많은 책이 쌓여 있지만 사용하기는 너무나 어려운 무용지물의 거대한 연구실이 탄생하게 된다. 우리가 말하고 있는 빅데이터 방대한 분량이지만 가치를 찾기는 쉽지 않은, 함부로 감당하기 어려운 무수한 데이터가 쌓여 있는 공간을 의미한다.

빅데이터를 유용하게 활용하기 위해서는 빅데이터에 대한 체계적인 관리가 선행되어야 한다. 기하급수적으로 늘어나는 데이터를 적절히 분류해서 관리하고 필요한 시점에 쉽게 찾을 수 있는 환경이 갖추어져 있어야 한다. 데이터가 아무리 많아도 어디에 존재하는지를 알 수 없다면 활용이 불가능하다. 또한 다양한 형태의 데이터를 적절히 결합하여 새로운 사실을 찾아낼 수 있어야 한다. 즉 데이터의 공유와 융합이 필요하다. 공유를 위해서는 서로를 인식하고 의미를 파악할 수 있어야 한다. 모두가 이해할 수 있고 상호 운용할 수 있는 방식을 따라야만 공유의 대상이 될 수 있다. 공유되는 데이터는 물론 적절한 품질을 만족해야 한다. 품질이 보장되지 않은 데이터가 공유되면 잘못된 결과를 전파하여 마치 바이러스와 같은 악영향을 미친다.

그러므로 데이터를 여러 가지 방법으로 공유하고 융합할 수 있으며 품질이 보장되는 거버넌스의 구축에 초점을 맞추어야 한다. 즉 올바른 빅데이터 거버넌스 체계가 정착되어야만 무엇인가를 만들 수 있고 지속적으로 부가가치를 창출해낼 수 있다.