IT & Future Strategy

빅데이터 시대의 인재, 데이터 사이언티스트의 역할과 가능성

제8호 (2012. 8. 8)

목 차

Ⅰ. 빅데이터 시대, 데이터 사이언티스트 확보 전쟁 / 1
Ⅱ. 데이터 사이언티스트의 역량과 조건 / 6
Ⅲ. 데이터 사이언티스트 육성을 위한 노력 / 11
Ⅳ. 결론 및 시사점 / 17

NIA 한국정보화진흥원

BIGdata 빅 데 이 터 전 략 연 구 센 터

'**IT&Future Strategy**'는 21세기 한국사회의 주요 패러다임 변화를 분석하고 이를 토대로 미래 정보사회의 주요 이슈를 전망, IT를 통한 해결방안을 모색하기 위해 NIA에서 기획, 발간하는 보고서입니다.

NIA의 승인 없이 본 보고서의 무단전재나 복제를 금하며, 내용에 대한 문의나 제안은 아래 연락처로 해 주시기 바랍니다.

▶ 발행인 : 김성태

▶ 작 성 : 한국정보화진흥원 국가정보화기획단 정보화전략연구부 | 빅데이터전략연구센터
권정은 선임연구원 (02-2131-0734, jekwon@nia.or.kr)

▶ 보고서 온라인 서비스
- www.nia.or.kr, www.itglobal.or.kr, www.bigdataforum.or.kr

요약

□ **세계는 빅데이터 시대의 패권을 쥐기 위한 데이터 사이언티스트 확보 전쟁중**

○ 빅데이터는 과학적 의사 결정으로 기업과 국가의 생산성 향상에 기여하는 혁신 도구로 각광

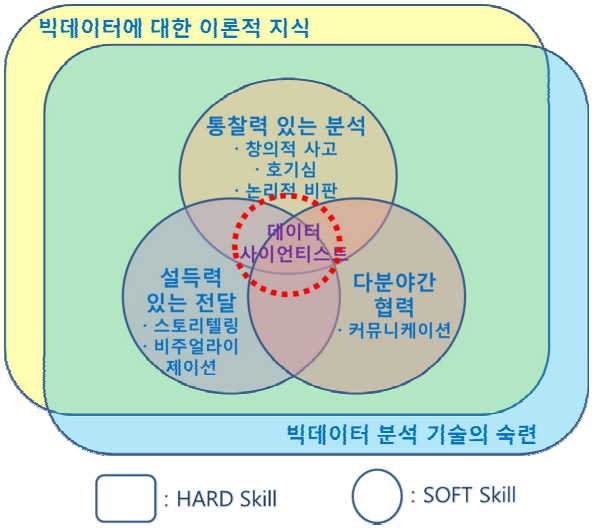
○ 빅데이터의 가능성 실현을 위해서는 **데이터 이면의 의미를 해석해내는 인재인 데이터 사이언티스트가 필수**

- 데이터 사이언티스트는 데이터의 다각적 분석을 통해 조직의 전략방향을 제시하는 기획자이자 전략가

○ 빅데이터 시대의 리딩 국가로의 도약을 위하여 **데이터 사이언티스트 육성을 위한 준비와 노력이 시급한 시점**

□ **우수한 데이터 사이언티스트란 빅데이터에 대한 이론적 지식과 분석 기술에 대한 숙련을 바탕으로 통찰력, 전달력, 협동 능력을 발휘할 수 있는 전문 인력**

< 데이터 사이언티스트의 역량과 조건 >



[HARD Skill]

- **① 빅데이터에 대한 이론적 지식**
  : 관련 기법에 대한 이해와 방법론 습득
- **② 분석 기술에 대한 숙련**
  : 최적의 분석 설계 및 노하우 축적

[SOFT Skill]

- **③ 통찰력 있는 분석** : 창의적 사고, 호기심, 논리적 비판
- **④ 설득력 있는 전달** : 스토리텔링, 비주얼라이제이션
- **⑤ 다분야간 협력** : 커뮤니케이션

□ **우수한 데이터 사이언티스트의 육성은 빅데이터의 국가적 성패를 좌우하는 핵심 전략**

○ 데이터 사이언티스트는 데이터를 통해 고부가가치를 창출하는 **지식기반사회의 고급인력이자 국가경쟁력 강화를 위한 인재**

○ 체계적이고 전략적인 데이터 사이언티스트 육성 전략을 위해서는 다음과 같은 준비가 필요

① 산-학-연-관 협력의 빅데이터 인력양성 체계 마련
- 국가 주도의 빅데이터 R&D 프로젝트 추진시, 다양한 분야의 참여주체간 협력체계를 구축

② 빅데이터 전담조직의 마련 및 위상 강화
- 빅데이터에 대한 지속적인 투자와 노력을 경주하기 위한 구조적 토대 마련
- 데이터 사이언티스트 육성 시장의 활성화를 지원

③ 빅데이터를 접할 수 있는 다양한 기회 제공
- 공공 데이터의 확대 공개를 통한 분석 역량 강화 지원
- 데이터 사이언티스트에 대한 사회적 관심도 제고

Ⅰ 빅데이터 시대, 데이터 사이언티스트 확보 전쟁

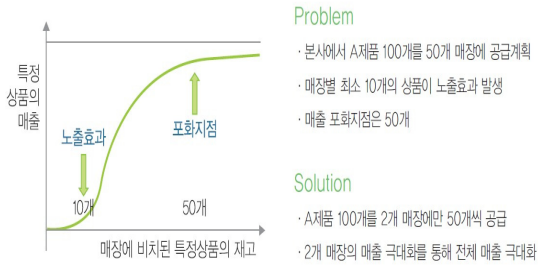
□ 빅데이터, 데이터 분석을 토대로 한 과학적 의사 결정으로 기업과 국가의 생산성 향상에 기여하는 혁신 도구로 각광

○ 기업체들은 자사의 **경영 전략에 빅데이터를 도입하여 수익 증대를 실현**

- ZARA[1]는 빅데이터를 활용하여 최대 매출 달성을 위한 재고 최적 분배 시스템을 개발함으로써 매출을 획기적으로 증가

  ※ Zara는 지난 5년간 연평균 17%에 달하는 매출 성장률 기록[2]

- Netflix[3]는 고객 맞춤형 추천 시스템을 개발, 고객 확보(560만→1,000만) 및 주요한 매출 발생지점(전체 매출의 약 60%)으로 활용[4]

Zara의 적정 재고 산출 알고리즘(예시)	Netflix의 고객 추천 시스템 : Cine-match
	

출처 : 한국정보화진흥원(2012. 5), '빅데이터로 진화하는 세상 - Big Data 글로벌 선진 사례'

1) 세계적 SPA(Speciality retailer of Private label Apparel) 브랜드
2) 조선일보(2012.8.7.), '자라(Zara) 얼마나 팔리기에...워런 버핏 제치고 세계 3위 부자'
3) 美 DVD 렌털 서비스 업체
4) 글로벌 IT 트렌드(2012. 3), '노무라연구소, 빅데이터 성공 열쇠는 리더십과 인재', (https://www.itglobal.or.kr/m_board/m_board_view.asp?seq=3228&root_code=30002&c_type=s&board_idx=2&)

○ 빅데이터를 **공공영역에 도입할 경우에도 높은 경제적 효과가 발생할 것**으로 예상

- McKinsey는 의료, 공공행정, 소매, 제조, 개인정보 부문에 빅데이터 적용시 1%의 추가 생산성 향상이 가능할 것으로 예측[5)]

  ※ 각 부문별로 $1,000억~$7,000억 규모의 경제적 효과 창출이 예상되며, 특히 연간 美 의료부문 $3,300억, 유럽 공공부문 €2,500억의 가치 예상

- 영국의 연구기관 Policy Exchange는 자국의 공공부문에 빅데이터 도입시, 연간 £160억~330억을 절감할 것으로 예측[6)]

  ※ 이는 영국 국민 1인당 연간 약 £250~500를 절감하는 효과로, 정부 총 예산인 £7,000억의 2.5~4.5%에 해당

- 우리 나라도 공공 분야의 빅데이터 활용 효과가 10.7조원에 달할 것으로 추산[7)]

빅데이터의 활용 가치

출처 : KT경제경영연구소(2012. 2), 'Big Data : 미래를 여는 비밀 열쇠' 재인용, McKinsey(2011. 5)

영국 공공부문 빅데이터의 잠재력

출처 : Policy Exchange(2012. 7), 'The Big Data Opportunity: Making government faster, smarter and more personal'

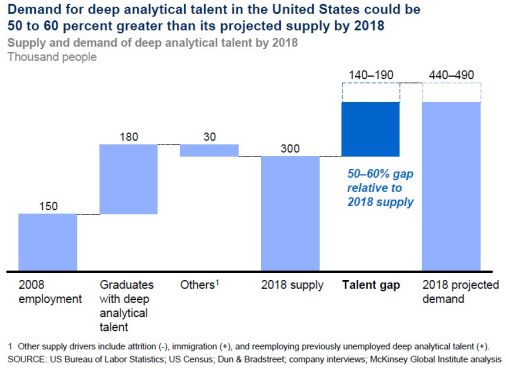
5) McKinsey Global Institute(2011. 5), 'Big data: The next frontier for innovation, competition, and productivity'
6) Policy Exchange(2012. 7), 'The Big Data Opportunity: Making government faster, smarter and more personal'
7) 국가정보화전략위원회(2011), '빅데이터를 활용한 스마트 정부 구현'

□ 빅데이터의 가능성을 실현하기 위해서는 우수한 역량을 갖춘 데이터 사이언티스트의 확보가 필수

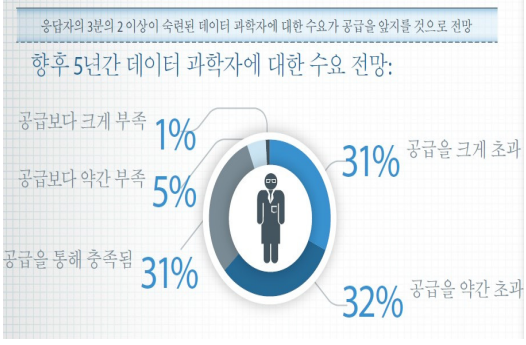
○ **빅데이터의 가치를 충분히 이끌어내기 위해서는** 데이터 이면의 의미를 해석해내는 인재, 즉 **데이터 사이언티스트가 필수**

- 데이터 사이언티스트는 데이터의 다각적 분석을 통해 **조직의 전략방향을 제시하는 기획자이자 전략가**
- 숙련된 데이터 사이언티스트는 단기간에 육성되기가 어렵고, 지속적인 현장 경험을 통해 전문 지식과 노하우를 축적하여 만들어지는 고급 인적자원
- 이러한 데이터 사이언티스트에 대한 수요는 계속해서 증가하고 있으며, 향후에도 이러한 추세는 이어질 것으로 전망

< 데이터 사이언티스트에 관한 수요 예측 >



출처 : McKinsey Global Institute(2011. 5), 'Big data: The next frontier for innovation, competition, and productivity'



출처 : EMC, '데이터 과학자 조사 (http://korea.emc.com/collateral/software/white-papers/data-scientist-infographic.pdf)

◆ 데이터 사이언티스트(Data Scientist)

: Wikipedia에 따르면, 데이터 사이언스(Data Science)는 데이터 엔지니어링, 과학적 방법론, 수학, 통계학, 고급 컴퓨팅(Advanced Computing), 비주얼라이제이션(Visualization), 해커(Hackers)적 사고방식, 영역별 전문지식을 종합한 학문이며, 이러한 수행이 가능한 者를 데이터 사이언티스트로 정의

○ 빅데이터의 가능성을 간파한 **글로벌 기업과 국가에서는 이미 데이터 사이언티스트 확보를 위한 적극적인 노력을 경주 중**

- eBay는 5천여 명 규모의 고객 데이터 분석인력, EMC는 데이터 사이언티스트로 구성된 애널리틱스 랩을 갖추는 등 각 기업별로 빅데이터 내부역량 강화에 심혈
- 미국, 일본, 싱가포르 등 주요 선진국에서도 빅데이터 기술 및 서비스 개발과 함께 데이터 사이언티스트 육성 정책을 적극적으로 추진중

< '누가 데이터 사이언티스트를 고용하는가?'...잘 나가는 IT기업이 원하는 者 >

'Network World'는 페이스북, 구글, 그루폰 등 최근 구인광고를 통해 데이터 사이언티스트를 채용하는 것으로 알려진 주요 IT기업의 구인 요건을 분석

공통적으로는 **이공계 분야의 석·박사학위**나 **충분한 현장 경험**, **다양한 분석 툴의 활용능력**을 바탕으로 **문제 해결의 강한 열정, 신기술의 학습 의욕, 리더십** 등을 필요 요건으로 제시

facebook
- 관련 분야 이공계 석사 또는 박사학위 소지자이거나 관련 업무 경력 4년 이상
- 정성적 접근을 통한 분석 및 문제 해결에 경험이 풍부한 사람
- 데이터와 관련한 고난이도의 질문에 답을 찾고 경험적 연구에 열정적으로 도전할 사람
- R, Matlab, SAS 등의 분석 툴 전문가

Google
- 관련 분야 이공계 석사 또는 박사학위 소지자이거나 관련 업무 경력 4년 이상
- 데이터에서 결론을 도출할 수 있는 사람
- 다른 사람에게 기술을 가르치고 새로운 기술 습득하는 것을 즐기는 사람
- 리더십을 갖추고 자기 주도적인 사람

GROUPON Collective Buying Power
- 데이터 마이닝, 기계 학습, 통계 분석, 응용 수학 박사 학위 또는 이와 동급의 자격 소지자
- SAS, R 등 분석 툴을 완벽하게 다룰 수 있는 사람

※ 출처 : CIO(2012. 6), '데이터 과학자를 고용하는 IT기업은 누구?' (http://www.ciokorea.com/slideshow/12832?slide=2#slideshow_anchor, 'Who's hiring data scientists? Facebook, Google, StumbleUpon and more') 재인용, Network World(2012. 6),

□ 빅데이터 시대의 리딩(leading) 국가가 되기 위해서는 우수한 데이터 사이언티스트 육성을 위한 준비와 노력이 시급

○ 빅데이터에서 新가치를 찾아내고 이를 사회 각 분야에 적용하기 위해서는 **양질의 전문인력을 충분히 확보하는 것이 최우선**

- 『분석으로 경쟁하라[8]』의 저자인 Thomas H. Davenport[9] 교수는 "빅데이터 경쟁에선 '사람'이 핵심"이라며, **인력 양성에 초점을 맞춘 빅데이터 정책 추진의 필요성** 강조[10]
- 국내 빅데이터 시장은 서서히 성장하고 있는 추세이나, 빅데이터 시장을 주도할 만큼의 전문인력 및 연구역량은 미흡[11]

○ 빅데이터 시대를 이끌 **우수 인력의 체계적이고 선제적인 육성 전략은 빅데이터 리딩 대한민국 실현에 초석**이 될 것

- 미국은 지난 3월, 'Big Data R&D Initiative'를 통해 빅데이터 인력양성에 대한 적극적 투자계획을 발표하는 등 주요 선진국에서는 이미 전문인력 양성을 위한 발빠른 움직임 시작
- 우리 나라도 빅데이터 R&D 및 인력양성에 중점을 둔 범부처 차원의 육성 전략 마련이 필요

< 오바마 정부 'Big Data R&D Initiative'의 인력 양성 계획 >

- 차세대 데이터 사이언티스트와 엔지니어를 육성을 위한 **학제간 교육 프로그램 개발 독려**
- 버클리 대학의 'Expeditions in Computing' 프로젝트에 $1,000만을 투자, **데이터의 정보 전환 기술**(기계학습, 클라우드 컴퓨팅, 크라우드 소싱)**을 통합 지원**
- 대학생을 대상으로 복합적 데이터의 그래픽 및 비주얼라이제이션 기술을 훈련시키는 **연구 트레이닝 그룹에 $200만의 상금 제공**

※ 출처 : 정지선(2012. 4), '성공적인 빅데이터 활용을 위한 3대 요소 : 자원, 기술, 인력', 한국정보화진흥원 재인용, 전자신문(2012. 3.30)

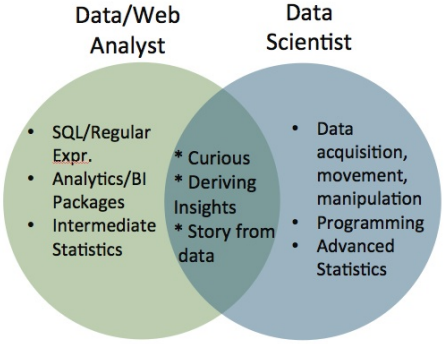
8) 'Competing on Analytics : The New Science of Winning', Harvard Business School Press(2007)
9) 美 Babson College 교수
10) '데이븐포트 교수 "빅데이터 경쟁에선 '사람'이 핵심, 조선일보, 2012. 4. 19.
11) '빅데이터 시대, 정책은 거북이 걸음', 파이낸셜 뉴스, 2012. 7. 11.

Ⅱ 데이터 사이언티스트의 역량과 조건

□ 우수한 데이터 사이언티스트는 다방면에 걸쳐 복합적이고 고도화된 지식과 능력을 갖추는 것이 필수

○ 데이터 사이언티스트는 **여러 분야에 걸친 전문성과 이를 복합적으로 활용하는 능력이 중요**

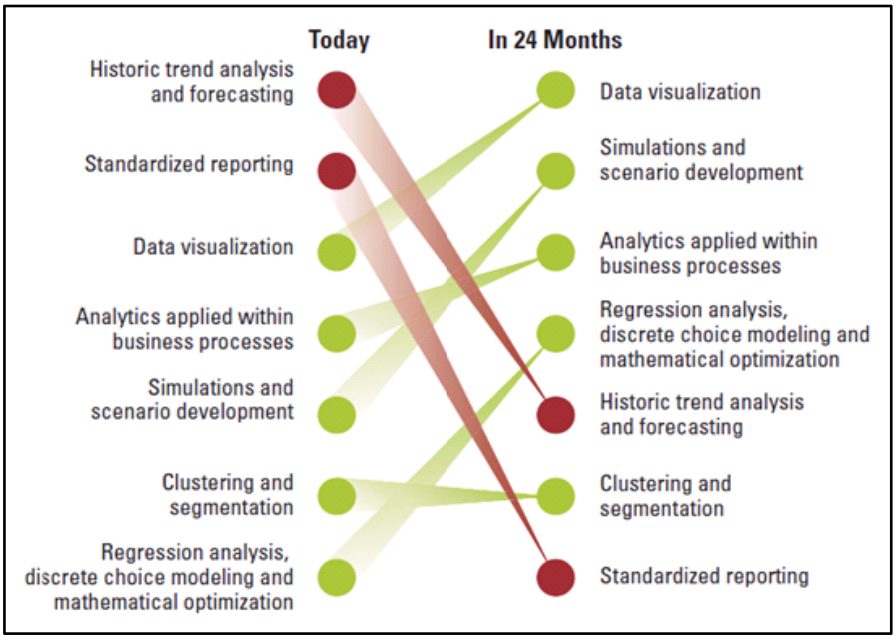
- 수학, 통계학, 컴퓨터공학 등 다양한 분야에 걸친 심도 있는 지식이 필수
- 이는 기존에 조직에서 데이터 분석을 수행하던 데이터 분석가(Data Analyst)와 비교하여 한층 높은 수준의 전문성과 다양성을 요구받는 것

데이터 사이언티스트의 역할을 정의하는 단어들	데이터/웹 분석가 vs. 데이터 사이언티스트
	
・데이터 사이언티스트는 분석(analytics), 다양한 경험(experience), 팀(team) 차원의 협력 등 그 **역할이 다양하고 복합적**	・데이터 사이언티스트는 데이터의 수집 및 처리에 대한 이해와 고급 수준의 통계적 지식 등 **기존에 비해 높은 역량을 요구**
출처 : 'Defining and Differentiating the Role of the Data Scientist' (http://blogs.gartner.com/doug-laney/defining-and-differentiating-the-role-of-the-data-scientist/)	출처 : 'Data/Web Analyst vs. Data Scientist' (http://blogs.splunk.com/2012/05/16/analytics-staffing-for-big-data/)

○ 특히 함의(Insight)를 도출하여 경영 전략을 수립하는 것이 빅데이터의 핵심 가치로 부각되면서, **복합적이고 고도화된 사고 능력에 대한 요구는 더욱 증대**

- 데이터 중심 관리자들이 빅데이터의 도입을 통해 향후 2년 내 달성할 것으로 기대하는 바를 살펴보면,
  - 과거의 추세 분석 및 예측이나 표준화된 보고와 같은 기초적인 데이터 분석 결과는 그 중요도가 하락하는 반면,
  - **데이터의 시각화 기술이나 시뮬레이션 및 시나리오 개발을 통한 비즈니스 프로세스에의 분석 적용**이 점차 중요한 기능으로 부각될 것으로 전망

< 데이터 중심 관리자(Data-driven Managers)들은 어느 곳을 향하고 있는가? >



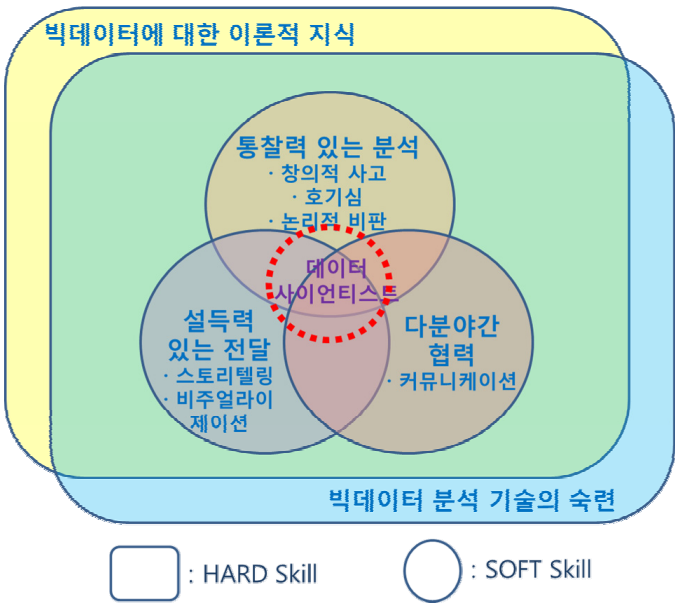
* 출처 : Steve LaValle etc.(2011), 'Big Data, Analytics and the Path From Insights to Value', MIT Slon Management Review, Winter)

< 전문가가 본 데이터 사이언티스트의 역량 >

전문가	제시한 역량
DJ Patil Greylock Partners社 데이터 사이언티스트, 'Building Data Science Teams' 저자	· **기술적 숙련도(Technical Expertise)** : 몇몇 과학 분야에 대한 전문적 지식 · **호기심(Curiosity)** : 내부에 숨겨진 것을 알고자 하는 욕구, 문제 해결을 위해 명확한 가설 집합을 만드는 능력 · **스토리텔링(Storytelling)** : 이야기를 전달하고 효과적으로 대화하기 위해 데이터를 활용하는 능력 · **영리함(Cleverness)** : 창의적 방식으로 문제를 다르게 보는 능력
Gartner	· **데이터 관리(Data Management)** : 데이터에 대한 이해 · **분석 모델링(Analytics Modeling)** : 분석론에 대한 지식 · **비즈니스 분석(Business Analysis)** : 비즈니스 요소에 초점 · **소프트 스킬(Soft Skill)** : 커뮤니케이션(Communication), 협력(Collaboration), 리더십(Leadership), 창의력(Creativity), 규율(Discipline), 열정(Passion)
Sunil Shirguppi LinkedIn, 인터내셔널 데이터 서비스 책임자	· 데이터 자체를 다루기 위한 능력 – **데이터 수집력(Data Gathering)** – **표준화(Standardization)** – **통계(Statistics)** – **모델링(Modeling)** · 데이터를 분석하여 의미 있는 결과를 도출하고 이를 적용하기 위한 능력 – **호기심(Curiosity)** – **직관력(Intuition)** – **비주얼라이제이션(Visualization)** – **커뮤니케이션(Communication)**
John Rauser Amazon, 수석 엔지니어	· **수학과 공학능력** : 데이터의 분석 및 이를 위한 솔루션 작동을 위해 필요 · 인문학적 소양 – **비판적 시각** : 가설 수립 및 검증에 필수 – **글쓰기 능력 및 대화 능력** : 분석 결과의 전달과 이해, 설득을 위해 필요 · **호기심 및 행복** : 통찰력의 발휘 및 일에 대한 만족과 보람도 필요

□ 우수한 데이터 사이언티스트는 빅데이터에 대한 이론적 지식과 분석 기술에 대한 숙련을 바탕으로 통찰력, 전달력, 협동 능력을 발휘할 수 있는 전문 인력

< 데이터 사이언티스트의 역량과 조건 >



○ **[HARD Skill] 빅데이터의 처리 및 분석에 필요한 이론적 지식 및 기술적 숙련**

- **① 빅데이터에 대한 이론적 지식**
  - 빅데이터 처리 프로세스에 대한 이해
  - 통계, 데이터마이닝, 기계학습, 자연어 처리 등 관련 기법에 대한 이해와 방법론 습득

- **② 분석 기술에 대한 숙련**
  - 데이터의 특성에 따라 적절한 분석 기술을 선택·적용하는 최적의 분석 설계
  - 다수의 다양한 빅데이터 처리 경험을 통한 노하우 축적

○ **[SOFT Skill] 데이터 속에 숨겨진 가치를 발견하고 새로운 발전 기회를 만들어내기 위한 능력**

- **③ 통찰력 있는 분석**
  - 창의적 사고 : 데이터가 나타내는 피상적 의미를 넘어 새로운 관점으로 조망. 개별 데이터 및 데이터세트(dataset)간의 연관성 발견
  - 호기심 : 문제를 해결하고자 하는 열망. 지속적으로 문제해결 과정을 탐색해가는 끈기
  - 논리적 비판 : 과학적 절차와 방법에 따라 데이터의 수집과 분석을 수행하고, 이 과정을 스스로 검토하며 오류를 수정

- **④ 설득력 있는 전달**
  - 스토리텔링 : 도출된 인사이트를 조직의 이해관계자와 의사결정자가 이해하고 업무 전략에 반영할 수 있도록 설득
  - 비주얼라이제이션 : 데이터의 시각화를 통한 효과적인 의미 전달. 인포그래픽(infographic)

- **⑤ 다분야간 협력**
  - 커뮤니케이션 : 서로 다른 지식을 지닌 다수의 전문가 간에 충분한 의견 교류와 정보 교환을 위해 필요한 능력

◆ a Superman vs. a Team for the great data science?

: 데이터 사이언티스트는 여러 분야에 걸쳐 깊은 지식과 기술을 보유한 뛰어난 개인이어야 하는가? 아니면 다양한 분야의 전문가들이 모여 논의를 통해 결론을 이끌어내는 조직의 형태여야 하는가?

많은 전문가들은, 데이터 사이언티스트에게 요구되는 조건과 그 깊이는 실제로 한 사람이 다 갖추는 것이 매우 어렵기 때문에, '위대한 한 사람'을 찾기보다 **다양한 역량을 지닌 사람을 모아 조직을 꾸리는 것을 추천**

팀 구성을 위해서는 다양한 배경을 가진 사람들간의 원활한 의사소통과 의견 교환을 위한 커뮤니케이션이 필수적

* 참고 : '데이터 과학, 혼자 할 수 있는 일이 아니다', 전자신문, 2012. 5. 24

Ⅲ 데이터 사이언티스트 육성을 위한 노력

Ⅰ. 학위과정 및 학제간 연구기관

□ Institute for Advanced Analytics

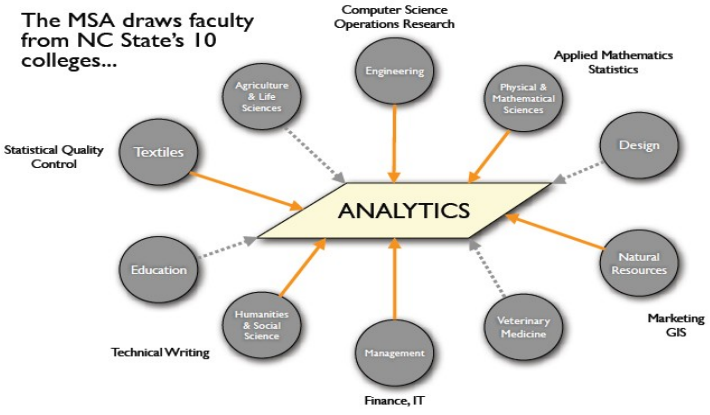
○ 美 North Carolina State Univ.에서 **데이터 사이언티스트 육성을 위해 운영중인 분석과학 석사**(Master of Science in Analytics)**과정**

- 3학기 10개월에 걸쳐 데이터 마이닝 및 최적화, 데이터 비주얼라이제이션, 금융 분석, 커뮤니케이션, 팀워크 능력 향상 등에 초점을 맞춰 진행되는 데이터 분석 능력 심화 프로그램
- 수학, 통계학, 공학, 경영학, 물리학 등 전공자 또는 MBA 출신을 중심으로 매년 80여 명의 신입생 선발하여 교육

○ 활발한 산학 협력 및 팀워크 중심 교육으로 **현장 경험이 풍부한 실전 인력 양성**

- 4~5명의 학생으로 구성된 팀 단위로 데이터 분석 트레이닝
- 이를 위해 기업체로부터 데이터 분석 연구계획서와 데이터를 모집하고, 프로젝트 형태로 각 팀이 실제 분석 수행

< MSA의 학제간 연구를 위한 교수진 구성 >



□ 충북대학교 비즈니스 데이터 융합학과

○ 지식경제부 지원으로 **국내 최초로 빅데이터 분야 소프트웨어 인력양성 석사과정** 신설(2012년~, 공학석사)

- 고용계약형 석사과정으로, 대학과 기업이 공동으로 신입생을 선발, 고용계약을 체결하며 졸업 후 해당 기업에서 의무복무
- 기업에서 수업의 50% 이상을 담당하고, 인턴쉽, 현장 실습 등을 포함하여 실제적 분석 능력 제고에 중점

○ **산학연계 및 다학제간 수업을 통한 실무전문가 양성**

- 디지털정보융합학과, 경영정보학과, 소프트웨어학과, 정보통계학과, 경영학부 관련 교수진 및 다수의 데이터 분석 관련 기업이 참여
- 산학협동 석사논문을 통한 산학공동기술개발, 기업 맞춤형 우수 SW 실무인재 확보 등 산학 연계의 장점 확보

< 충북대 비즈니스 데이터 융합학과 교과과정 >

대학원 교과과정

구분	필수	선택
기초공통	대용량 Database	빅데이터 EDA 분산병렬처리 기업프로세스 통합분석 대용량 멀티미디어 자료처리
전공	빅데이터 세미나 I	비즈니스 데이터 분석 정보검색과 활용 IT 산업과 빅데이터 컴퓨팅 R-데이터마이닝 비즈니스 인텔리전스 기업정보시스템 구축 I 기업정보시스템 구축 II ABAP 프로그래밍 클라우드 시스템 빅데이터 세미나 II
일반		인턴쉽 I 인턴쉽 II 인턴쉽 III
연구		연구과제

선수과목

참이교수 소속학과 학부과정 과목 중 개인 특성에 맞게 지정

디지털정보융합학과
경영정보학과
경영학과/정보통계학과
컴퓨터 관련학과

개인별 취약점 보완

교과과정 기본원칙

※ 한국BI데이터마이닝 학회는 고교생을 대상으로 데이터마이닝에 대한 이해를 높이고 다양한 실제 사례를 경험하는 단기 캠프를 연 2회 개최

□ 서울대학교 데이터과학과 지식창출 연구센터

○ 데이터과학과 관련된 다양한 학문분야의 교수진들이 협력하여 공학, 의학 등 타 분야에의 접목과 응용방안을 연구

- 데이터의 분석(통계학), 통계를 용이하게 하는 시각화(컴퓨터공학), 이를 처리하고 가공하는 데이터마이닝(산업공학) 등 다분야간 융합을 위해 설립(2011. 12월 개소)
- 데이터 과학의 이론과 모델에 대한 연구를 주로 추진하고 있으며, 현재 의학 분야에서의 데이터 관련 연구를 추진중

< 서울대 데이터과학과 지식창출 연구센터 연구목표 >

데이터 과학의 다학제적 학문 체계 정립	- 다양한 데이터 분석 및 지식창출을 위한 '이론·기술·응용'의 전 단계를 통합하는 데이터 과학의 신 학문영역 체계 정립 - 데이터 수집, 변환, 추출, 저장, 시각화, 처리, 계산 기술의 집대성
데이터 과학의 융합적 연구 모형 개발	- 데이터 과학의 분야별 전문 지식의 연계를 바탕으로 '투입·가공·창출' 단계의 유기적 모형 개발 - 연관분야와의 상호작용과 공동연구를 위한 비약적 발전 성취
세계를 선도하는 데이터 과학연구 및 브레인 풀 구축	- 연계지식창출과 지식파급을 위해 학문분야간, 산업분야간의 국제적 전문가 교류 활성화 - 분야간 연계를 통한 국제적 연구 인력풀 구축 및 연구분야 선도
데이터 활용의 선도적 응용 연구	- 다양한 산업영역의 방대한 데이터를 실제 활용하여 데이터 과학에 대한 소프트웨어와 기법들을 연구 - 연구자, 산업 관계자들에 있어서 선도적 테스트베드 역할

< 현재 추진중인 연구과제 >

구분	연구과제명	세부과제명
제1총괄과제	데이터 모델링 이론 연구	[1-1] 확장된 우도 예측법과 고차원 데이터분석
		[1-2] 시공간 데이터 분석
제2총괄과제	데이터 테크놀로지	[2-1] 효율적인 미래 데이터 분석을 지원하는 애자일 데이터 웨어하우스 연구
		[2-2] 데이터 분석 및 지식 창출을 위한 일반적인 시각화 및 인터액션 모델 연구
제3총괄과제	지식 발견 및 생성	[3-1] 사회/기술 데이터로부터의 지식발견 기법 연구
		[3-2] 바이오/의학 데이터로부터의 지식발견 기법 연구

Ⅱ. 데이터 사이언티스트 단기과정 및 이벤트

□ Data Science Summer Institute(DSSI)

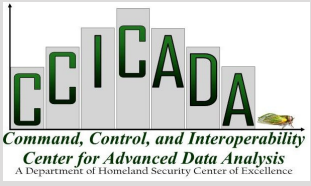
○ 美 Univ. of Illinois에서 진행하는 6주 기간의 데이터 사이언스 교육과정

- 기초 및 응용 수학, 데이터 사이언스 관련 토론 수업, 그룹 리서치 등으로 진행
- 컴퓨터 과학 전공자 및 대학원 신입생, 충분한 수학 및 프로그래밍 경험자 중 미국 거주자를 대상으로 대상자 선발

○ 과학 연구, 교육, 정부업무 분야의 **인적자원 개발을 위하여 국가가 기금을 지원**

- DSSI를 주관하는 Multimodal Information Access & Synthesis Center는 美 국토안보부로부터 기금을 지원
- 이는 미래의 국가발전을 위한 인력양성의 일환으로 추진되는 것으로, 국토안보부는 Univ. of Illinois를 비롯한 다수의 대학 및 연구소와 협력관계를 맺고 고급 데이터 분석연구를 지원

◆ Command, Control, and Interoperability Center for Advanced Data Analysis(CCICADA)



- 美 국토안보부(Department of Homeland Security)에서 **데이터 분석을 통한 국가 안보 강화를 목적으로 전문인력 양성을 지원**하는 프로그램
- 주요 대학 및 연구소와 협력관계를 맺고 기금을 지원함으로써 빅데이터 연구개발 및 전문인력 양성을 지원

□ 국내외 분석업체 주관 교육 및 컨퍼런스

○ EMC는 전문인력 양성을 위한 **EMC Academic Alliance**와 분석방안과 사례 공유의 장인 **Data Scientist Summit**을 개최

- EAA는 세계 각국의 대학 및 교육기관을 대상으로 **데이터 분석교육을 지원하는 산학협력 프로그램**

  ※ EAA 가입 기관은 정보 인프라스트럭처 교육 교재 및 실습 환경을 지원받으며, 현재 우리 나라도 숭실대, KAIST 등 4개 대학이 가입중

- Summit에서는 빅데이터 관련 전문가들이 모여 **분석 전략 및 기술 발전을 논의**하고 **네트워킹의 장으로 활용**

  ※ 올해의 주제는 '예측, 변환, 그리고 데이터 드림팀 구현'

○ SAS는 자사의 분석 툴을 활용한 **데이터 분석 교육과정**과 빅데이터 분석 및 적용기법을 소개하는 **Analytics Conference**를 개최

- SAS School: 데이터 분석의 기초 툴인 SAS 프로그래밍의 습득과 이를 활용한 분석 실습으로 이루어진 단기 프로그램
- SAS Advanced Analytics European Tour: 유럽 각국의 SAS Institute에서 운영되는 교육프로그램을 다양하게 경험
- Conference는 예측 모델링, 경영과학, 비즈니스 분석 등을 주제로 진행되며, **3일짜리 사전 트레이닝 코스를 통해 실습 기회 제공**

  ※ 올해는 데이터 마이닝, 텍스트 마이닝, 데이터 비주얼라이제이션, 데이터 최적화, 소셜 미디어 분석 등의 다양한 토픽으로 진행

○ 국내 소셜분석업체 다음소프트는 **Opinion Mining Workshop**을 개최하여 소셜 빅데이터의 분석 방법론과 활용 방안을 공유

- **소셜미디어를 활용한 트렌드 분석 방법론 및 실제 적용사례**를 논의하는 장으로 연1회 개최
- 국내 분석업체의 기술력과 실제 적용 사례의 탐색이 가능

□ Kaggle : We're making data science a sport

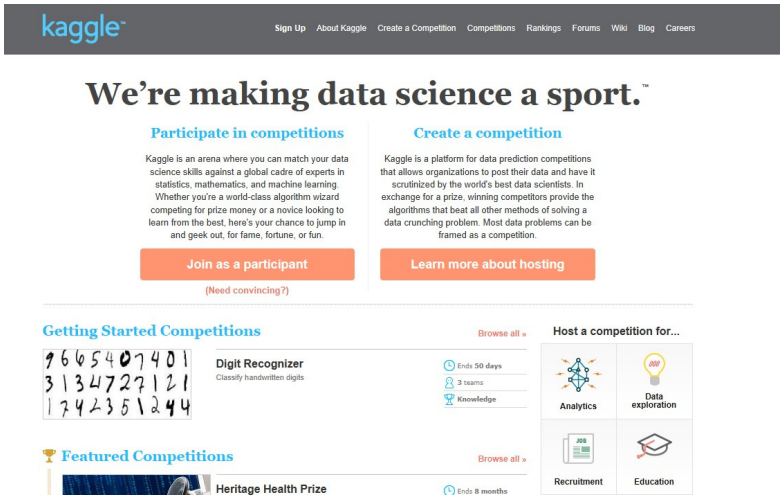
○ 기업의 실제 데이터를 분석하여 그 결과에 따라 포상하는 **데이터 사이언티스트 경진대회 중개사이트**

- 자사의 데이터를 분석하여 숨겨진 가치를 발견하려는 기업과 데이터 분석의 실전 경험을 쌓고자 하는 데이터 사이언티스트들이 참여
  - · 참여 기업 : 자사의 데이터 분석 및 가치 창출 업무를 아웃소싱
  - · 참여자 : 다수의 데이터 분석 경험 및 노하우 축적
- 참가자/팀에게는 실제 기업의 빅데이터가 제공되며, 이를 가장 잘 처리해 좋은 결과를 낸 참가자/팀에게 상금을 지급

○ **실제 데이터를 활용한 실전 경험을 제공**함으로써, 우수한 역량을 갖춘 **데이터 사이언티스트의 발굴과 지속 훈련이 가능**

- 대회의 우승자/팀은 자신의 분석 기법을 공개하게 되고, 이를 통해 우수 분석 기법 및 노하우가 공유
- 각 참가자는 우승 횟수에 따라 랭킹이 매겨져 순위가 공개되며, 데이터 사이언티스트의 구인(求人)이나 참여자간 네트워킹도 가능

< Kaggle : 데이터와 분석 능력의 만남의 장 >



Ⅳ 결론 및 시사점

□ **빅데이터 시대를 제패하기 위해서는 빅데이터 속에 숨겨진 가치를 찾아낼 '사람'에 대한 준비가 최우선**

○ 빅데이터 시대의 주도권 확보를 위한 각국의 뜨거운 승부에서 **승패를 가를 히든 카드는 가치를 발견해 낼 '사람'**

- 많은 자원(빅데이터)과 이를 처리할 도구(분석 기술)가 있어도, 그 안에 숨겨진 가치를 이끌어내는 핵심 요소는 사람(인력)
- 같은 데이터와 분석 기술을 가지고도 얼마나 창조적 시각으로 유의미한 결과를 도출해내는가가 빅데이터의 성패를 좌우

○ 데이터 사이언티스트는 데이터 속에서 고부가가치를 창출하는 **지식기반사회의 고급인력이자 국가경쟁력 강화를 위한 인재**

□ **체계적으로 준비된 인재육성 전략으로 빅데이터 리딩 국가로서의 확고한 기반을 마련해야 할 시점**

○ 빅데이터는 IT 강국의 위상 한계에 봉착한 오늘의 상황을 극복하고 **또 한 번의 혁신적 도약을 위한 기회**

- 우리 나라의 IT 시장은 이미 글로벌 기업의 기술 잠식과 주요 분야의 전문인력 부족으로 경쟁력을 위협받고 있는 상황
- 그러나 세계 최고 수준의 IT 인프라, 스마트 기기에서 축적되는 데이터의 양 등 빅데이터 선점에 유리한 환경은 이미 확보

○ **데이터 사이언티스트 육성을 위한 체계적 전략 수립과 정책 추진**은 빅데이터 시대 선도국가로의 도약을 위한 원동력이 될 것

- 철저한 준비와 노력으로 빅데이터 선진 국가로서의 기반 강화를 위해 노력

□ 빅데이터 전쟁의 패권을 좌우할 핵심 인재인 데이터 사이언티스트 육성을 위해서는 다음과 같은 준비가 필요

1 산-학-연-관 협력의 빅데이터 인력양성 체계 마련

○ 빅데이터 인력양성의 핵심은 관련 이론과 경험을 겸비한 **실무형 인재의 육성**

- 사회과학, 자연과학, 인문학 등 다방면에 걸친 지식과 기술의 연계 학습과 실제 분석 경험을 통한 역량의 강화가 중요
- 따라서 산-학-연-관의 다양한 주체가 협력, 이론과 실습을 병행함으로써 실무형 인재의 육성을 위한 전략을 수립

○ 국가 주도의 빅데이터 R&D 프로젝트 추진시, **다양한 분야의 참여주체간 협력체계를 구축**하여 **프로젝트 수행을 통해 전문인력을 양성하는 '두 마리 토끼' 전략**이 필요

- 빅데이터 기술 및 서비스 개발 프로젝트는 그 자체로 연구개발인 동시에 빅데이터에 대한 이해를 높이고 관련 경험을 확장하는 인력 육성의 기회로 활용 가능
- 또한 빅데이터 분석은 특정한 분야의 지식이나 기술에 한정되지 않으므로, **관련 분야와의 학제간 공동연구 형태로 진행하는 것이 바람직**
- 빅데이터 R&D를 추진하고 있는 주요국에서는 산학이 협력하는 전문인력 양성 및 다학제간 연구를 장려
  - 미국, 일본 등은 정부의 R&D 프로젝트 발주시 대학 및 연구기관과의 공동 프로젝트로 수행하는 경우가 다수[12)]
  - 또한 지구과학, 수학, 공학, 의학, 생물학 등의 관련 학문과의 다학제간 연구개발을 적극 지원

12), 14) 정보통신산업진흥원(2012. 3), '美·日의 Big Data R&D 전략과 우리나라의 대응과제', IT R&D 정책동향(2012-3)

2 빅데이터 전담조직의 마련 및 위상 강화

○ 빅데이터에서의 가치 발견을 위해서는 **빅데이터 전담조직의 존재 및 위상 강화**가 중요

- 빅데이터의 가치는 데이터의 수집부터 분석까지 단선적인 과정을 거쳐 항상 도출되는 것이 아니며, 때로는 끊임없는 시행 착오와 오류를 거쳐야 할 수도 있음
- 이러한 과정에서 지속적인 투자와 노력을 기울이기 위해서는 힘있는 빅데이터 전담조직의 뒷받침이 필수

○ 힘있는 **빅데이터 전담조직은 데이터 사이언티스트 육성 시장의 활성화에 영향**을 미쳐, 우수한 인재 확보에 기여

○ 빅데이터에 적극적인 선진 국가 및 기업은 **별도의 전담조직이나 책임자를 지정하고 빅데이터 분석에 박차**

- 주요 선진국에서는 빅데이터 전담기구를 마련하고, 빅데이터 발전을 위한 국가전략 수립 및 다부처 협력의 정책추진 시도

◆ **세계 주요국의 빅데이터 정책 추진 기구**

- **미국 :** 대통령 직속기관인 과학기술정책실(Office of Science and Technology Policy)의 주도로 “Big Data Research & Development Initiative”를 발표(2012. 3)
- **영국 :** 부처의 공공정보 공개에 대한 정책적 조언을 제공하는 데이터 전략위원회(Data Strategy Board)설립(2012. 3)

- 일부 기업에서도 최고 데이터 책임자(Chief Data Officer) 직제를 신설하고 빅데이터 전략 추진을 전폭적으로 지원

3 빅데이터를 접할 수 있는 다양한 기회 제공

○ 데이터 사이언티스트로서의 숙련을 위해서는 실제로 **빅데이터를 처리·분석해 본 경험과 그를 통한 노하우 축적이 필수**

- 데이터 사이언티스트를 구인하는 기업들의 주요한 채용조건 중 하나가 다년간의 분석 경력일 정도로 데이터 사이언티스트에게 분석의 경험과 노하우 습득은 중요

○ 다양한 데이터로 분석 경험을 쌓을 수 있도록 **공공 데이터를 확대 공개하여, 숨겨진 가치를 찾을 수 있도록 지원할 필요**

- 기업체는 데이터를 자사가 보유한 '숨겨진 무기'로 인식하고 내부 기밀로 다루기 때문에 쉽게 확보하여 분석해보기 어려움
- **정부 및 공공기관이 보유한 고품질의 신뢰도 높은 데이터를 공개하여 고품질의 신뢰도 높은 데이터에 대한 접근성을 강화**

◆ **세계 주요국의 공공데이터 공개 정책 추진 현황**

- **미국 :** 매우 강력한 데이터 공개 정책을 추진하고 있는 오바마 정부는 data.gov를 통해 약 45만 여 개의 데이터세트를 공개
- **영국 :** data.gov.uk를 통해 약 9천개의 데이터세트를 공개하고, 정부의 데이터 공개 정책을 제시하는 '오픈 데이터 백서'를 발간(2012. 6)

○ 또한 빅데이터와 관련된 다양한 행사를 개최함으로써 **데이터 사이언티스트에 대한 사회적 관심도 제고**

- 빅데이터 분석 역량 컨테스트, 데이터 사이언스 컨퍼런스 및 관계자 간 네트워킹의 장 등을 마련
- 대학 진학 청소년, 관련 분야 종사자 등에게 빅데이터 관련 학과로의 진학이나 커리어 패스 설계에 대한 관심을 유도

참고 자료

[1] 국가정보화전략위원회(2011. 11), '빅데이터를 활용한 스마트 정부 구현'

[2] 정보통신산업진흥원(2012. 3), '美·日의 Big Data R&D 전략과 우리 나라의 대응과제'

[3] 정지선(2012. 4), '성공적인 빅데이터 활용을 위한 3대 요소 : 자원, 기술, 인력', 한국정보화진흥원

[4] 한국정보화진흥원(2012. 5), '빅데이터로 진화하는 세상 - Big Data 글로벌 선진 사례'

[5] KT경제경영연구소(2012. 2), 'Big Data: 미래를 여는 비밀 열쇠'

[6] DJ Patil etc.(2011), 'Building Data SCience Teams', O'Reilly

[7] Gartner(2012. 3), 'Emerging Role of the Data SCientist and the Art of Data Science'

[8] McKinsey Global Institute(2011.5), 'Big data: The next frontier for innovation, competition, and productivity'

[9] Steve LaValle etc.(2011, Winter), 'Big Data, Analytics and the Path From Insights to Value', MIT Slon Management Review

[10] 글로벌IT트렌드(2012. 3. 22), '노무라연구소, 빅데이터 성공 열쇠는 리더십과 인재' (https://www.itglobal.or.kr/m_board/m_board_view.asp?seq=3228&root_code=30002&c_type=s&board_idx=2&)

[11] 디지털타임스(2012. 3. 22.), '[알아봅시다] 데이터 사이언티스트'

[12] 전자신문(2012. 5. 24.), '데이터 과학, 혼자 할 수 있는 일이 아니다',

[13] 조선일보(2012. 4. 19.),'데이븐포트 교수 "빅데이터 경쟁에선 '사람'이 핵심"'

[14] 조선일보(2012.8.7.), '자라(Zara) 얼마나 팔리기에...워런 버핏 제치고 세계 3위 부자'

[15] 파이낸셜 뉴스(2012. 7. 11.),'빅데이터 시대, 정책은 거북이 걸음'

[16] CIO(2012.6), '데이터 과학자를 고용하는 IT기업은 누구?' (http://www.ciokorea.com/slideshow/12832?slide=2#slideshow_anchor, 'Who's hiring data scientists? Facebook, Google, StumbleUpon and more')

[17] Doug Laney, 'Defining and Differentiating the Role of the Data Scientist' (http://blogs.gartner.com/doug-laney/defining-and-differentiating-the-role-of-the-data-scientist/)

[18] EMC, '데이터 과학자 조사' (http://korea.emc.com/collateral/software/white-papers/data-scientist-infographic.pdf)

[19] Forbes(2011. 10. 7.), 'Amazon's John Rauser on "What Is a Data Scientist?"'

[20] Rahul Deshmukh, 'Data/Web Analyst vs. Data Scientist' (http://blogs.splunk.com/2012/05/16/analytics-staffing-for-big-data/)

2012년도 IT & Future Strategy 보고서

- 제1호(2012. 3), 「실리콘밸리가 전망하는 IT분야 현재와 미래」
- 제2호(2012. 4), 「선진국의 데이터기반 국가미래전략추진 현황과 시사점」
- 제3호(2012. 4), 「성공적인 빅데이터 활용을 위한 3대 요소 : 자원, 기술, 인력」
- 제4호(2012. 4), 「빅데이터 시대 : 에코시스템을 둘러싼 시장경쟁과 전략분석」
- 제5호(2012. 5), 「빅데이터 시대의 데이터 자원 확보와 품질관리 방안」
- 제6호(2012. 6), 「빅데이터 시대 : 공공부문 빅데이터 추진방향」
- 제7호(2012. 7), 「빅데이터 시대, AI의 새로운 의미와 가치」
- 제8호(2012. 8), 「빅데이터 시대의 인재, 데이터 사이언티스트의 역할과 가능성」

1. 본 보고서는 정보통신진흥기금으로 수행한 정보통신연구개발사업의 연구결과입니다.
2. 본 보고서의 내용을 발표할 때는 반드시 행정안전부 정보통신연구개발사업의 연구결과임을 밝혀야 합니다.
3. 본 보고서의 내용은 한국정보화진흥원의 공식 견해와 다를 수 있습니다.
4. 본 보고서 내용에 대해 무단전재를 금하며, 가공 · 인용할 때는 반드시 「한국정보화진흥원」이라고 출처를 밝혀 주시기 바랍니다.