

IT & Future Strategy

빅데이터 시대, AI의 새로운 의미와 가치

제7호 (2012. 7. 25)

목 차

- I. 빅데이터의 도전 / 1
 - II. AI와 빅데이터 / 6
 - III. AI를 활용한 빅데이터 사례 / 12
 - IV. 결론 및 시사점 / 18
-

‘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
오정연 책임연구원 (02-2131-0726, oh.jy@nia.or.kr)

▶ 보고서 온라인 서비스

– www.nia.or.kr, www.itglobal.or.kr, www.bigdataforum.or.kr

□ 빅데이터의 도전: 빅데이터를 가치있게 하는 +a가 필요

- 빅데이터에 대한 무지개 빛 전망과 함께 이를 우려하는 부정적인 시각들 등장
 - 전례 없이 빠른 속도로 팽창하는 데이터의 양에 비해 이를 저장, 분석 및 해석하는 기술엔 한계가 존재
 - ※ 美정부 IT 네트워크인 'Meritalk'는 빅데이터의 가능성(possibility)과 현실(reality)에는 아직까지 'Gap'이 존재한다고 분석
 - 빅데이터의 등장과 함께 모델을 만들고 이를 검증하는 방식의 전통적인 과학 연구방법론이 퇴색하고 있다는 주장도 등장
 - ※ Wired의 편집장 Chris Anderson은 자료의 홍수 속에서 더 이상 기존의 과학적 방법론은 의미가 없다고 주장
- 빅데이터의 진정한 가치는 커다란 데이터 자체에 있는 것이 아니라 추출할 수 있는 새로운 사회적 가치에 있음
 - ⇒ 더 이상 똑같은 통계 분석방법과 기존의 분석 도구가 아닌 새로운 접근방식으로의 전환이 필요

□ 인간을 닮은 기술 AI, 빅데이터를 만나 수면위로 부상

- AI는 컴퓨터에 인간과 같은 지능을 실현하기 위한 시도 및 일련의 기술을 의미
 - 그동안 현실성이 떨어지는 기술로 터부시됐던 AI가 빅데이터를 만나면서 가능성에 대한 기대 증폭
 - AI의 분석력 및 예측과 빅데이터가 만나면 AI의 신뢰성 및 현실 가능성이 증폭
 - ※ 기계학습, 에이전트, 자연어처리, 패턴인식 등의 AI기술은 빅데이터 처리를 위한 핵심 기술이자 기반
- ⇒ 인간의 학습능력과 추론능력, 지각능력, 이해능력 등을 실현하는 AI는 빅데이터의 궁극적인 미래가 될 수 있음

□ 빅데이터와 AI, 상호보완효과를 통한 시너지 창출

- 이미 선진국 및 선진기업들은 빅데이터와 AI를 접목한 다양한 서비스를 제공 중

<AI와 빅데이터 접목 사례>

구글 독감예보 서비스	<ul style="list-style-type: none"> · 독감, 인플루엔자 등 독감과 관련된 검색어 쿼리의 빈도를 조사, '구글 독감 동향'이라는 독감 확산 조기 경보체계 마련 · 미국 보건당국보다 한발 앞서 시간 및 지역별 독감 유행정보 제공
인공지능슈퍼컴퓨터 왓슨	<ul style="list-style-type: none"> · IBM은 인공지능 슈퍼컴퓨터 '왓슨(Watson)'을 개발 · 인간의 언어를 이해하는 것은 물론 각종 분야에 대한 해박한 지식, 게임전략 등을 습득 · 방대한 양의 데이터를 기반으로 스스로 학습하고 적응 · 의료계와 금융계 등 다양한 분야에 활용 중
위치주적서비스 MetaCarta	<ul style="list-style-type: none"> · 수천 개의 서류를 분석하고 그 결과를 대규모 지리 데이터베이스를 이용하여 상호 체크 검증하도록 하여 위치를 정확히 제시 · 범죄활동의 패턴을 추적하여 그 활동패턴이 집중되는 위치도 찾아낼 수 있음
애플 Siri 음성인식서비스	<ul style="list-style-type: none"> · 인공지능을 강화한 음성인식 기술 · 사용자의 위치정보와 언어를 받아들여 문맥을 이해하고 사용자가 원하는 검색결과나 기능을 실행
밀라노 교통정보 시스템	<ul style="list-style-type: none"> · 교통흐름에 영향을 주는 다양한 정보들을 종합적으로 분석하여 최적의 교통안내 서비스를 제공 · 5분에서 15분 간격으로 수집된 데이터를 분석하여 향후 2~24시간을 예측 가능
구글 자동번역시스템	<ul style="list-style-type: none"> · 기존의 방법과 달리 통계적 기법을 활용한 실시간 자동 번역시스템을 개발 · 번역시스템이 각 언어의 일반적인 표현을 스스로 학습

□ 빅데이터의 가능성을 극대화시키고 새로운 가치를 끌어내기 위해서는 AI기술과의 접목이 필수

- 빅데이터와 AI를 결합한 다양한 비즈니스 모델 개발
- AI기술 및 인력의 전략적 육성
- 기술 + 사회문화적 요소에 대한 면밀한 검토

I

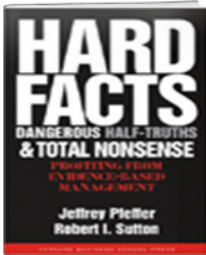
빅데이터의 도전

□ 새로운 혁신 패러다임, ‘증거기반경영’

- 최근 광범위하고 **정확한 데이터와 과학적 방법론에 의해 도출된 증거**에 따른(evidence-based) 기업경영과 국가정책 수행의 중요성 부각
 - 스탠퍼드 석좌교수 제프리 페퍼는 자신의 저서 **‘Hard Facts’**를 통해 **‘최상의 증거를 토대로 한 의사결정’**의 필요성을 주장
 - 증거기반경영이란 미약한 통념, 관행, 경험, 벤치마킹 대신 과학적으로 입증된 증거에 입각해 의사결정을 내리는 경영방식
- 최근 이슈가 되고 있는 ‘빅데이터’는 증거기반경영을 뒷받침하기 위한 중요한 자원으로 인식
 - **증거 데이터가 많을수록 이론은 더욱 정교화** 되고 탄탄해질 수 있음
 - ※ 통계학적으로 봤을 때 표본이 많아지면 표본오차가 작아짐
- 이러한 맥락 아래 ‘빅데이터’가 IT업계 최대의 화두로 떠오르면서, 기업은 물론 공공부문까지 빅데이터에 대한 투자 활발
 - 빅데이터 처리능력이 기업 또는 국민경제의 생산성 향상 및 경쟁력 강화로 이어질 수 있다는 **‘데이터 산업혁명’** 주장도 대두

- 데이터는 화폐 또는 금처럼 새로운 경제적 자산이 될 것 -다보스포럼
- 빅데이터는 미래 경쟁력을 좌우하는 21세기 원유다! - 가트너-
- 빅데이터는 비즈니스의 지형을 바꿀 10가지 기술 트렌드- 맥킨지 컨설팅

[증거기반경영 성공사례]



Hard Facts, Dangerous Half-Truths, and Total Nonsense: Profiting from Evidence-based Management

Jeffrey Pfeffer, Robert I. Sutton 저

HARVARD BUSINESS

2006.03.20

◇ 세계 최대의 카지노 호텔 그룹 '해라스 엔터테인먼트' 사례

- 해라스는 토털 리워드(Total Reward)'라는 고객 로열티 프로그램에 입각해 거대한 데이터베이스와 실시간 분석시스템을 구축
- 고객들의 다양한 행동패턴 - 방문일시, 주로 즐기는 게임, 지출금액, 게임시간, 좋아하는 식사, 음료 등을 철저히 분석해 의사결정을 내림
- 2000년대 초반 150만명의 고객데이터 분석을 통해 진짜 충성고객을 선별하여 관리함으로써 업계 1위로 도약
- ※ 진정한 충성고객이 부자 여행객, 도박 중독자가 아니라 카지노 근교에 사는 평범한 50대~60대 직장인이었음을 밝혀내고, 기존의 호화시설, 호화쇼 대신 세심한 고객 서비스에 주력하는 방식으로 마케팅을 전환

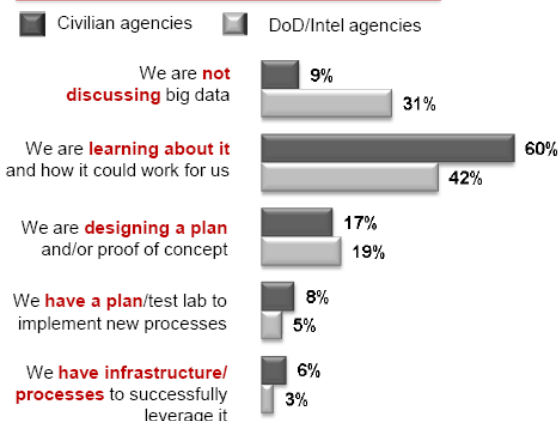
◇ 기후변화 대응이 세계 공통의 최우선 정책과제가 된 과정

- 대중들에게 생소했던 기후변화나 온실가스 감축이 초미의 정책과제로 부상한 이유는 강력한 경제적, 과학적 증거가 제시되었기 때문
- 증거1: 2006년 10월 영국 재무성이 발간한 Stern 보고서
 - ※ 기후변화 추세를 방치하면 2020년 이후 대처비용이 전 세계 GDP의 20%에 육박할 것을 경고
- 증거2: 2007년 11월 발간된 IPCC 4차 보고서
 - ※ 과거 200여년간 탄소배출량과 지구 온도 추이를 분석해 지구온난화 주범이 인류가 배출하는 온실가스임을 과학적으로 증명

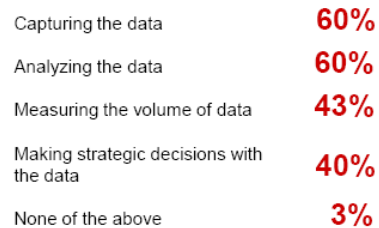
□ 빅데이터 갭(Gap) : Promise vs Capabilities

- 빅데이터에 대한 무지개 빛 전망과 함께 이를 우려하는 부정적인 시각들도 등장
 - 빅데이터의 진정한 가치는 커다란 데이터 자체에 있는 것이 아니라 추출할 수 있는 새로운 사회적 가치에 있음
 - 그러나 전례 없이 빠른 속도로 팽창하는 데이터의 양에 비해 이를 저장, 분석 및 해석하는 기술엔 한계가 존재
 - 빅데이터를 축적하는 행위와 분석하는 행위로 나누어 볼 때 현실적 측면에서 현재의 붐은 호들갑스러운 측면이 있다는 의견
- 美정부 IT 네트워크인 'Meritalk'는 빅데이터의 가능성(possibility)과 현실(reality)에는 아직까지 'Gap'이 존재한다고 분석
 - ※ 151명 연방 정부 CIO 및 IT관리자 대상 'Big Data Gap' 조사(2012.3) 실시
 - 대부분의 기관들이 빅데이터의 혜택에 대해 동의하고 있는 것에 반해 실질적인 기술이나 어플리케이션 등은 초기 수준
 - ※ 실질적으로 현재 데이터를 제대로 활용하고 있는 기관도 적으며, 데이터의 소유권(ownership) 문제 조차도 확립되지 않은 것으로 나타남
 - Meritalk는 적어도 3년은 지나야 빅데이터가 가져오는 실질적인 혜택에 대해 논의할 수 있을 것으로 분석

Where is your agency with big data?



What is your agency currently doing with the large amounts of data it collects?*



□ 이론의 종말: Half-Truths, Total Nonsense

- 빅데이터의 등장과 함께 모델을 만들고 이를 검증하는 방식의 전통적인 과학 연구방법론이 퇴색하고 있다는 주장도 등장
 - Wired의 편집장 Chris Anderson은 자료의 홍수 속에서 더 이상 기존의 과학적 방법론은 의미가 없다고 주장
 - 통계학자 George Box의 '모든 모형은 틀렸고 일부만 쓸만하다.'는 'Half-Truths 논의'가 빅데이터 시대에 부합한다는 것
- ※ 구글의 연구책임자인 Peter Norvig은 더 나아가 "모든 모형은 틀렸고 (Total Nonsense) 앞으로 이런 모형 없이 성공할 수 있다."라고 단언

<The End of Science> (Wired Magazine: Issue 16.07)

- 데이터의 양이 적은 경우에는 일부 데이터를 보고 적절한 가설을 세우고, 나머지 데이터를 가지고 검증이 가능
- 그러나 테라, 페타바이트 이상의 데이터를 가지고 가설을 세우고 이 가설의 참/거짓 여부를 판별하는 것이 사실상 불가능
- ※ 페타바이트 수준에선, 정보는 3,4차원 정도로 단순한 분류와 순서의 문제가 아니라 불가지 차원의 통계적 문제로 완전히 새로운 접근이 필요

- 인식의 한계치를 넘어선 데이터: 팩트가 아닌 패턴을 발견하려는 노력이 필요¹⁾
 - 예전의 데이터 분석의 대상이 미시적 대상이나 가시적 대상의 범주였다면, 빅데이터는 가시적일 수는 있으나 인식의 한계치를 넘어선 데이터 범주를 의미
 - 빅데이터 분석의 요체는 거대함 속에서 드러나는 하나의 흐름을 파악하는 것이 중요
- ※ 데이터를 모니터하여 평상시와는 다른 특이한 아웃라이어를 잡아내는 것이 빅데이터 분석의 핵심

1) 데이터 분석가가 바라본 빅데이터 현상과 대응을 위한 준비방법, 고태영 이노지에스 고문

□ 빅데이터의 새로운 가능성, AI에 주목

- 빅데이터를 활용한 증거기반 경영이 주목받고 있는 가운데, 데이터를 분석하고 해석하는 능력이 핵심 경쟁력으로 부상
 - 그러나 더 이상 똑같은 통계 분석방법과 기존의 분석 도구가 아닌 새로운 접근방식으로서의 전환이 필요한 시점
 - 빅데이터 시대에는 스스로 학습을 통해 맥락(패턴)을 구성하는 AI적 접근이 필요
 - ※ 빅데이터를 가장 잘 활용하고 있는 구글사의 임직원 상당수가 AI 전문가이며, 구글 서비스의 대부분은 AI에 기초하고 있음
- 구글은 최근 빅데이터를 활용하여 자가학습이 가능한 인공 신경망을 개발, AI와 빅데이터간의 가능성을 제시
 - 구글X 연구소는 1만6천개에 이르는 컴퓨터 중앙처리장치(CPU) 코어와 10억 건 이상의 데이터 연결을 처리하는 모델을 도입
 - ※ 인공신경망은 따로 ‘고양이’라는 단어를 입력해 학습시키지 않았음에도 불구하고, 유튜브 동영상 이미지에서 고양이를 스스로 구분해냄
 - 사실상 대규모 분산 컴퓨팅 인프라가 사람의 뇌 역할을 할 수 있음을 밝힌 셈이며, 빅데이터의 미래가 인공지능 분야로 연결될 수 있음을 보여줌
- AI와 빅데이터간의 결합이 활발한 가운데, 우리나라의 AI 분야는 고착상태로 경쟁력이 낮고 업계 활력도 낮은 상황
 - 한국이 전략적으로 빅데이터 분야를 선도해 나가기 위해서는 AI부문에 대한 현실파악과 철저한 준비가 필요
 - 이에 본 보고서는 AI와 빅데이터와의 관계를 재조명하고, 한국 빅데이터 발전을 위한 정책방향을 제시

II

AI와 빅데이터

□ 인간을 닮은 기술, AI

- 사전적인 의미의 ‘인공지능(Artificial Intelligence)’은 철학적인 개념으로써 인간이나 지성을 갖춘 존재 또는 시스템에 의해 만들어진 인공적인 지능을 의미
 - 인공지능이라는 용어는 ‘56년 영국 다트머스회의에서 존 매커시에 (John McCarthy)의해 처음 사용
 - ※ 일반적으로 컴퓨터에 인간과 같은 지능을 실현하기 위한 시도 및 일련의 기술을 의미
 - 컴퓨터과학이라는 관점에서 “수리적 모델을 이용하여 지적 능력을 연구하는 학문” (Charniak and McDermott, 1985) 이라는 정의가 일반적으로 통용

< AI의 역사 >

▶ 1950년대: AI의 새벽녘

- 기계에 의한 계산이 가능해지고 컴퓨터가 개발됨에 따라 철학, 수학, 논리학, 심리학 등의 분야에서 인간의 지적활동을 행하는 기계에 대한 논의가 시작
- ‘56년 다트머스회의에서 J.McCarthy에 의해 인공지능이라는 용어가 시작

▶ 1960년대: Good Old Fashioned AI

- 인공지능 기법을 이용하여 실세계문제에 적용하는 전문가시스템에 관한 연구가 활발히 진행

▶ 1970년대: 현실로부터의 반격

- AI의 현실 가능성에 대한 반격이 심했던 시기로, 대규모 문제나 복잡한 문제에 있어서의 AI의 적용이 어렵다는 의견 분분
- 다양한 테크닉과 알고리즘들이 개발

▶ 1980년대: AI의 산업화

- 상업적 데이터베이스 시스템 개발이 시작
- 각 국에서 AI연구에 대한 보조 및 투자가 활발해지기 시작

▶ 1990년대~ : 현재에서 미래로

- 직관이 아닌 엄밀한 이론과 견고한 실험을 통해 현실세계의 문제를 대상으로 AI연구 추진

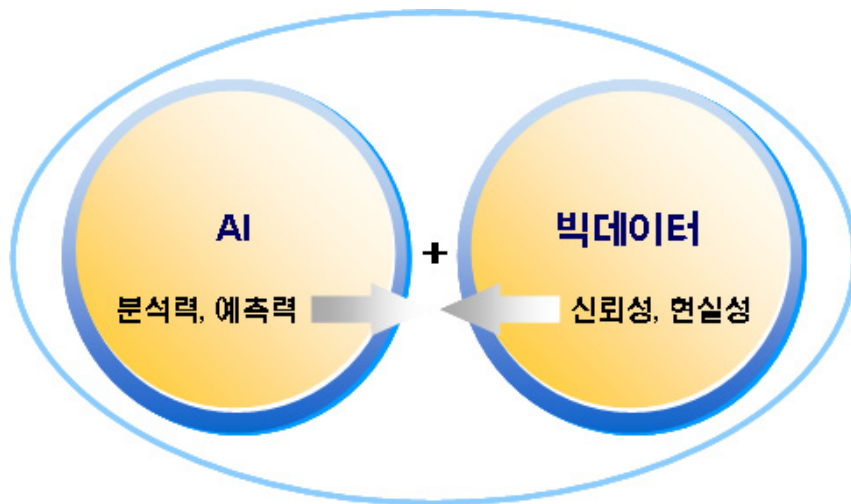
출처: <http://www.ai-gakkai.or.jp/jsai/whatsai/AIhistory.html>

□ AI, 데이터 분석 및 처리를 위한 핵심 기술기반

- 현재 인공지능 분야에서의 관심분야는 매우 광범위하며, AI는 데이터 분석 및 처리를 위한 핵심 기술기반

관련 기술분야	주요내용
패턴인식 (Pattern recognition)	<ul style="list-style-type: none"> · 기계에 의하여 도형·문자·음성 등을 식별시키는 것 · 현재로서는 제한된 분야에서 실용화되고 있고, 본격적인 패턴 인식은 아직 연구단계
자연어처리 (Natural language processing)	<ul style="list-style-type: none"> · 인간이 보통 쓰는 언어를 컴퓨터에 인식시켜서 처리하는 일 · 정보검색·질의응답 시스템·자동번역 및 통역 등이 포함
자동제어 Automatic Control	<ul style="list-style-type: none"> · 제어 대상에 미리 설정한 목표 값과 검출된 되먹임(feedback) 신호를 비교하여 그 오차를 자동적으로 조정하는 제어
로봇틱스(Robotics) 인지로봇공학 (Automatic Control)	<ul style="list-style-type: none"> · 로봇에 관한 과학이자 기술학으로 로봇의 설계, 제조, 응용분야를 다룸 · 인지로봇공학은 제한된 계산자원을 사용해 복잡한 환경의 복잡한 목표를 달성하도록 하는 인식능력을 로봇에게 부여하는 기술
컴퓨터비전 (Computer vision)	<ul style="list-style-type: none"> · 컴퓨터 비전은 로봇의 눈을 만드는 연구분야로 컴퓨터가 실세계 정보를 취득하는 모든 과정을 다룸
기계학습 (Machine Learning)	<ul style="list-style-type: none"> · 새로운 정보를 학습하고, 습득한 정보를 효율적으로 사용할 수 있는 능력과 결부시키는 방법 · 작업을 반복적으로 수행함으로써 결과를 얻어내는 기술의 개선 과정
양자컴퓨터 Quantum computer	<ul style="list-style-type: none"> · 양자역학의 원리에 따라 작동되는 미래형 첨단 컴퓨터 · 양자역학의 특징을 살려 병렬처리가 가능해지면 기존의 방식으로 해결할 수 없었던 다양한 문제를 해결 가능
자동추론 (Automated Reasoning)	<ul style="list-style-type: none"> · 계산기과학의 한 분야로 추론의 다양한 측면을 이해함으로써 컴퓨터에 의한 완전한 자동추론을 가능하게 하는 소프트웨어 개발을 목표로 함 · 인공지능연구의 일부로 이론계산기과학 및 철학과도 깊은 관계가 있음
사이버네틱스 (Cybernetics)	<ul style="list-style-type: none"> · 생물 및 기계를 포함하는 계(系)에서 제어와 통신 문제를 종합적으로 연구하는 학문
데이터마이닝 (Data mining)	<ul style="list-style-type: none"> · 많은 데이터 가운데 숨겨져 있는 유용한 상관관계를 발견하여, 미래에 실행 가능한 정보를 추출해 내고 의사 결정에 이용하는 과정
지능엔진 (Intelligent Agent)	<ul style="list-style-type: none"> · 인공지능적 기능을 가진 소프트웨어 엔진 · 사용자를 보조하고 반복된 컴퓨터 관련 업무를 인간을 대신하여 실시하는 엔진
시맨틱웹 (semantic web)	<ul style="list-style-type: none"> · 컴퓨터가 정보자원의 뜻을 이해하고, 논리적 추론까지 할 수 있는 차세대 지능형 웹

[빅데이터와 AI의 관계]



<AI와 빅데이터의 상호보완효과>

- AI와 빅데이터가 만나면 서로의 약점을 보완하면서 윈윈효과를 창출 할 수 있음
 - AI분야는 마치 앞이 안 보이는 외길을 걷는 분야와 같다는 평가를 받아 오면서 실현가능성에 대한 의문이 지속적으로 제기되어 옴
 - 그러나 어마어마한 양의 빅데이터가 등장하면서 새로운 국면을 맞이하고 있음
 - 방대한 양의 데이터가 AI의 실현가능성 및 연구의 신뢰성을 높여주고 있는 것
 - ※ 방대한 양의 데이터 습득과 학습, 분석, 추론 등을 통해 알고리즘의 정확성 및 정교성을 높일 수 있음
 - 빅데이터 분야 역시 AI의 분석력, 추론력, 예측력 등을 활용하면 더욱 지능화된 가치있는 서비스 창출이 가능
- AI와 빅데이터는 상호보완적인 역할을 수행하면서, 서로에게 시너지효과를 만들어낼 수 있음

□ AI, 빅데이터 기반의 맞춤형서비스 제공을 위한 핵심기반

- 지난 수십년 동안 우리는 IT를 통해 개개인에 대한 맞춤형 서비스를 제공하기 위한 노력을 추진해 옴
 - 빅데이터의 등장은 이러한 개인별, 상황별 맞춤서비스 제공을 가능하게 하는 원천자원이며 AI는 이를 활용하기 위한 핵심 툴
 - 다양하고 방대한 정보들이 복잡하게 얽혀있는 현재의 상황에서는 동적인 상황에서 경험적으로 문제를 해결하는 AI 기술의 적용이 필수
- ※ AI는 공학, 바이오, 문헌 정보 등의 연구 결과를 지난 30년간 분석해 왔으며, 일반 사용자 데이터분야까지 확대됨에 따라 더욱 다양한 서비스 발굴 가능

<일반적인 컴퓨터 SW공학과 AI의 차이점>

- 컴퓨터 SW공학, "number cruncher"의 개념으로 시작
 - 고성능의 CPU와 메모리를 가진 기계를 이용하여 사람이 할 수 없는 속도의 빠른 연산을 수행하는 소프트웨어 기술로부터 시작
- AI, "symbol manipulator"의 개념을 보유
 - AI는 제한된 능력이지만 사람이 가지고 있는 추론, 사고, 계획, 학습의 기능을 구현하는 방안을 연구

- 빅데이터 활용이 기존의 DMS(Data Management System)나 진화된 CRM(Customer Relationship management)를 넘어서기 위해서는 AI의 결합이 필수
 - 현재까지 우리나라의 경우 진정한 빅데이터 활용 사례는 매우 드물며, 기존의 DBM나 CRM을 조금 확장 시킨 수준
 - 진정한 빅데이터 활용을 위해서는 기존의 분석방법이나 툴이 아닌 새로운 접근과 시도가 필요
 - AI분야의 추론, 학습, 계획 및 에이전트 기능 등의 창의적 활용을 통해 새로운 가능성을 찾는 노력이 필수

□ 빅데이터의 궁극의 미래는 AI를 완성하는 것?

- 사람의 뇌는 데이터나 정보를 인식한 뒤 이를 전기 신호로 보내 서로 소통하는 방식(폴 리버 노스웨스턴大 교수)
 - 사람의 뇌는 약 2.5페타바이트(PB)에 이르는 정보를 저장하고 처리
 - 방대한 데이터 수집을 아파치 Hadoop과 같은 데이터 분석 기술이 담당하면 AI기술도 새로운 국면을 맞을 수 있음
 - ※ 대규모 분산 컴퓨팅 인프라가 사람의 뇌의 뉴런과 같은 역할을 수행할 수 있다는 것
- 최근 구글X의 ‘인공신경망 연구’²⁾가 발표되면서 빅데이터와 AI 구현의 경계가 모호해지고 있음³⁾
 - 대규모 분산 컴퓨팅 인프라를 사용해 방대한 데이터를 학습한 인공지능 등장에 대한 기대 고조
 - 빅데이터 분석 플랫폼 업체인 호튼웍스 개발진은 ‘공각기동대’⁴⁾를 언급하며 빅데이터의 최종 목표가 인공지능 구현에 있다고 밝힘

<애니메이션 ‘공각기동대’가 보여주는 빅데이터의 미래상(호튼웍스)>

- 사람의 생각을 바탕으로 움직이는 기계, 개인 정보 중심의 각종 맞춤형 서비스, 사람처럼 생각하고 사람의 마음을 이해하는 안드로이드의 등장 모두 빅데이터로 구현 가능한 사회
- 대규모 데이터 분석을 통해 사람과 비슷한 생각을 하는 인공지능을 완성하는 것이 빅데이터의 미래

- 인간의 학습능력과 추론능력, 지각능력, 이해능력 등을 실현하는 AI는 빅데이터의 궁극적인 미래가 될 수 있음

2) 대규모 분산 컴퓨팅 인프라를 사용해 자가학습이 가능한 인공 신경망 연구

3) ‘인공지능’, 빅데이터의 미래인가, 이지영, 2012.7.3

4) 공각기동대: 1995년 일본에서 시로우 마사무네의 만화를 원작으로 하여 제작한 애니메이션 영화

[SF애니메이션 '공각기동대']



공각기동대
(攻殻機動隊: Ghost In The Shell, 1995)

SF판타지 애니메이션(일본)

2002년 4월12일 개봉

감독: 오시이 마모루

◇ 줄거리5)

- '거대 정보 네트워크 시대의 정체성'이라는 다소 무거운 주제를 가지고 있음
- 서기 2029년 뉴포트(New Port:새로운 항구)라는 일컬어지는 아시아의 한 국가가 배경이며, 고속의 광대한 정보 네트워크로 연결된 변혁의 시기
- 인간은 사이보그화된 기계육체를 이용하여 육체의 한계를 벗어나게 되었고, 네트워크를 이용한 각종 범죄가 성행
- 정보가 가장 큰 가치로 떠오르게 되면서, 고스트 해킹이 공공연하게 벌어지게 되자 외무부는 프로그램 2501을 개발하여 각국의 네트워크를 통해 전뇌(전자두뇌)로 침입하여 고스트 해킹을 준비
- 이 과정에서 프로그램인 인형사가 자의식을 가지게 되면서 생명을 요구하게 되는데, 이를 해결하기 위해 투입된 주인공인 쿠사나기 모토코와 융합

◇ 주요 AI관련 등장요소

- 시각적으로 보이지 않게 되는 장비인 '광학 미체'
- 인공지능을 가지고 전투를 수행할 수 있는 다각(多脚) 전차 '타치코마'
- 케이블을 직접 연결해 정신이 전자 정보를 받아들일 수 있게 개조된 뇌인 '전뇌'
- 전뇌를 통해 조종할 수 있으며 통상적인 신체보다 훨씬 더 강력한 힘이나 세밀한 제어가 가능한 기계 몸인 '의체'

5) 공각기동대 [攻殻機動隊], 네이버 백과사전

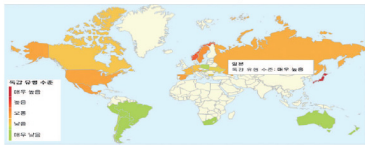
III

AI를 활용한 빅데이터 사례6)

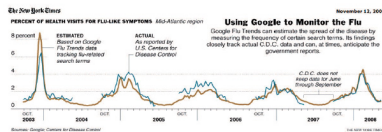
□ Google의 독감예보 서비스 및 경기동향지수

- 구글은 실시간으로 무수히 누적되는 검색어 분석을 통한 독감예보 서비스를 제공 중

구글 독감 동향(Google Flu Trends)



구글 독감 동향 및 실제 확산의 비교 결과



- 구글 홈페이지에서 독감, 인플루엔자 등 독감과 관련된 검색어 쿼리의 빈도를 조사, '구글 독감 동향(Google Flu Trends)'이라는 독감 확산 조기 정보체계 마련

※ 간단한 AI기술을 통해 독감 증세 환자가 늘면 '감기'와 관련된 단어를 검색하는 빈도가 함께 증가한다는 패턴을 발견해 내는 것

- 미국 보건당국보다 한발 앞서 시간 및 지역별 독감 유행정보 제공
 - 미국 질병통제 예방센터의 데이터 비교 결과, 검색 빈도 및 실제 독감증세를 보인 환자 숫자 사이에 매우 밀접한 상관관계 확인
- 또한 구글은 전 세계 온라인쇼핑 데이터로 경기동향지수를 개발7)
 - 구글은 전세계에서 취합한 방대한 온라인쇼핑 데이터를 기반으로 구글의 독자적인 경기동향 지수인 GPI(Google Price Index)를 산출
 - 웹과 오프라인에서 판매되는 상품이나 환경이 다르기 때문에 당장은 정부 통계인 소비자물가지수(CPI)를 대체할 수 없으나 경제관리 통계를 실시간성으로 집계할 수 있다는 장점

6) 빅데이터로 진화하는 세상, '빅데이터 글로벌 선진사례', 한국정보화진흥원, 2012

7) Financial Times(2010), 'Google to map inflation using web data'

□ ‘사람을 이긴’ 인공지능 슈퍼컴퓨터 왓슨

- IBM은 설립자 토마스 왓슨의 이름을 딴 인공지능 슈퍼컴퓨터 ‘왓슨(Watson)’을 개발



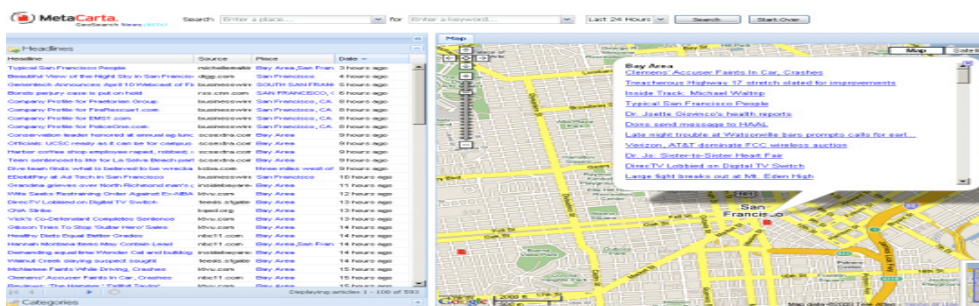
- 왓슨은 인간의 복잡하고 미묘한 자연어를 이해하기 시작한 첫 번째 컴퓨터
- ※ 말장난, 비꼬기 등을 비롯한 인간의 언어를 이해하는 것은 물론 각종 분야에 대한 해박한 지식, 게임전략 등을 습득
- 왓슨은 2011년 2월, 퀴즈쇼 ‘제퍼디(Jeopardy!)’ 최종전에 출전하여 압도적인 결과 차이로 승리
- 왓슨이 사람을 이길 수 있었던 힘은 인공지능에 기반한 데이터 분석능력
 - 3년간 수학, 과학, 인문학 등 다양한 분야에 대한 상식과 100만권의 책에 해당하는 지식을 축적했으며, 2억 페이지의 콘텐츠 보유
 - 15 테라바이트의 메모리와 2,880개의 프로세스 코어, 초당 80조의 연산능력을 보유하고 있으며, 정보의 상관관계를 따져 스스로 학습하고 상황에 적응 가능

<인공지능 슈퍼컴퓨터 ‘왓슨’, 의료계와 금융계에 적용>

- 미국 병원 ‘세튼 헬스케어 패밀리’가 환자 데이터 분석을 위해 왓슨의 ‘의료용 콘텐츠 및 예측 분석’ 기술을 채택(‘11.10월)
 - 방대한 양의 환자 데이터에서 임상 정보를 추출·분석해 미래를 예측하고 대비
 - 환자의 예방이 가능한 재입원과 병원 방문 횟수를 줄이는데 중점
- 미국 블룸버그통신은 인간의 언어를 이해하고 지식을 습득하는 왓슨이 월가의 금융 서비스에 활용될 계획이라고 보도(‘12.3월)
- IBM은 왓슨이 영역을 계속 확대해 2015년쯤 약 25억5000만달러의 추가 수입을 거둘 것으로 예상

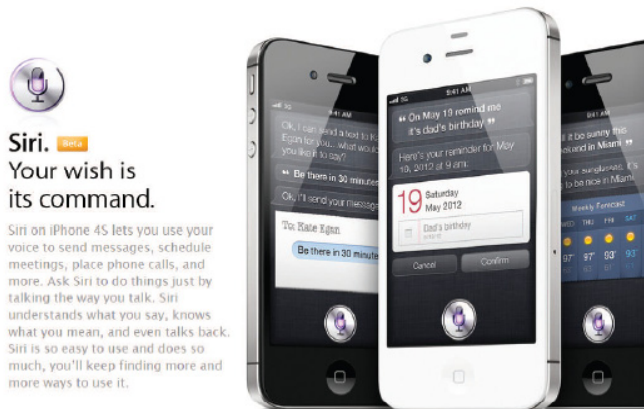
□ AI기술을 이용한 ‘위치 추적 서비스’⁸⁾

- 수배 중인 사람의 이름이나 관련 키워드를 입력하면 그 사람의 소재위치를 정확하게 지도에 맵핑(Mapping)해주는 프로그램이 개발
 - 미국 국방성, 중앙정보국, 국토보안부 등은 이러한 위치기반 정보 기술을 이용한 새로운 탐색작전으로 전환 중
 - 미래에는 범인 추적뿐 아니라 집이나 사무실에서도 누구나 찾고자하는 사람의 위치를 손쉽게 찾을 수 있게 될 것
- 이 프로그램은 메타카타(MetaCarta)라 불리는데 수천 개의 서류를 분석하고 그 결과를 대규모 지리 데이터베이스를 이용하여 상호 체크 검증하도록 하여 위치를 정확히 제시
 - 지금까지 이 ‘MetaCarta’ 프로그램은 주로 중앙정보국요원들이 테러전쟁과 같은 작전을 수행하기 위해 수집한 방대한 정보를 분석하는데 사용
 - 이 프로그램은 자동적으로 이메일이나 웹 페이지 같은 텍스트 문서로부터 지리적 연관성을 추출
 - ※ 어느 곳의 웹페이지 문서인지 이메일이 어디서 발신되었는지를 추적
 - 원하는 사람의 이름을 검색하면 지도상에 찾고자 하는 사람의 위치가 나타나며, 범죄활동의 패턴을 추적하여 그 활동패턴이 집중되는 위치도 찾아낼 수 있음



8) <http://demo.metacarta.com>

□ 애플의 똑똑한 음성인식 서비스 Siri



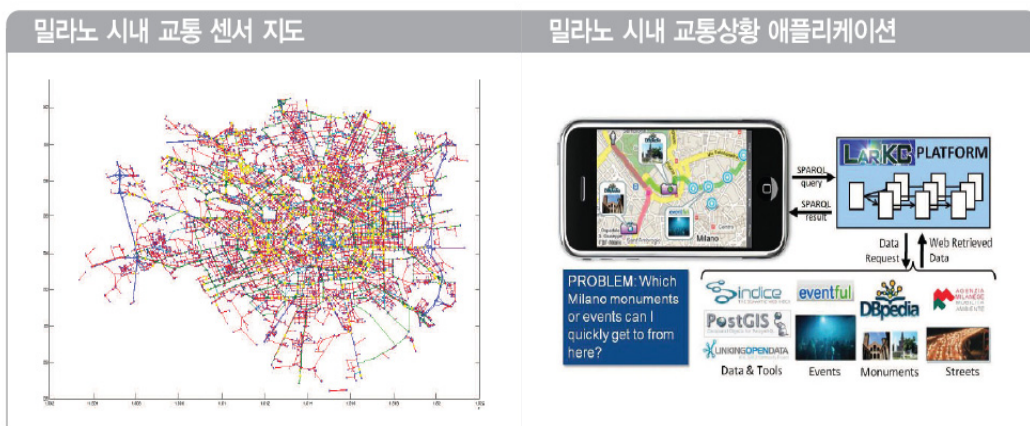
- 애플은 인공지능을 강화한 음성인식 기술을 이용하여 사용자가 원하는 검색결과나 기능을 실행해주는 서비스 개시
 - ※ Siri(Speech Interpretation and Recognition Interface)
 - Siri는 단순한 음성인식을 넘어 자연어의 뉘앙스까지 이해하는 똑똑한 음성인식 서비스로, '스티브 잡스가 진짜 그 안에 있다 (Steve is really inside)'는 우스개 소리가 나올 정도
 - 사용자가 질문을 하면 아이폰이 질문의 뜻을 자동적으로 파악하고 그에 맞는 답을 찾아주기 위해 애플 서버에 남아있는 데이터와 기존에 기록된 사용자의 정보를 최적화하여 제공
 - ※ 사용자의 위치정보와 언어를 받아들여 문맥을 이해하고 사용자가 원하는 검색결과나 기능을 실행

<음성인식 기술의 발전>

- 1950~1960년대: 최초의 음성인식 시스템은 오직 숫자만 인식 가능했으며, 1952년 벨 연구소가 최초로 음성인식 시스템 'Audrey' 개발
- 1970년대: 미국방부 주도로 음성인식 연구가 추진되었으며, 카메기멜론대는 1,011개의 단어를 이해할 수 있는 3세 어휘수준의 'Harpy' 개발
- 1980년대: 1,000개 단어를 인식하고 5,000개 단어의 어휘지원이 가능한 'Kurzweil Text-to-Speech' 프로그램 개발

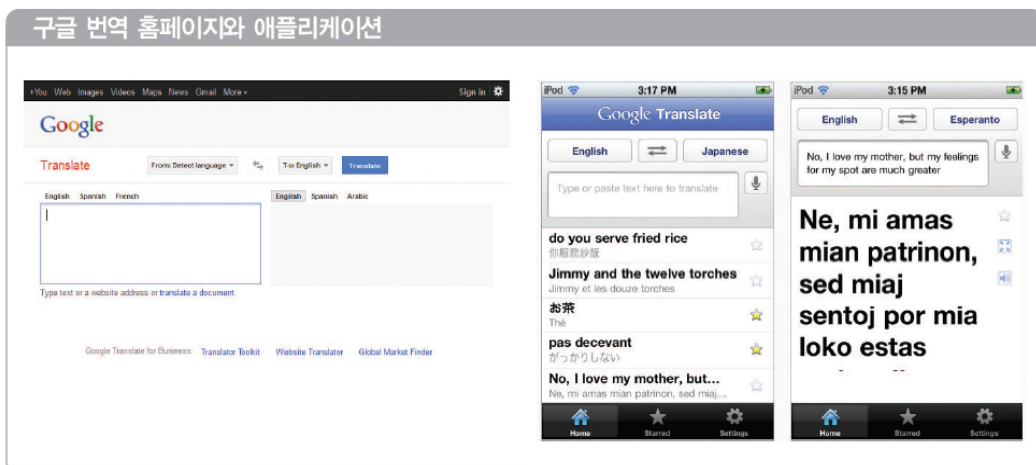
□ 밀라노의 지능형 교통정보 시스템

- 밀라노시는 교통흐름에 영향을 주는 다양한 정보들을 종합적으로 분석하여 최적의 교통안내 서비스를 제공 중
 - 경찰청의 사건정보, 기상청의 날씨 정보, 도로교통 상황, 주변 건물 및 도로공사 상황, 시위, 행사 등의 데이터 수집
 - ※ 5분 간격으로 교통 흐름량과 속도를 측정하여 데이터화하였으며, 데이터 분석을 위해 900여개의 센서를 도로 곳곳에 부착
 - ※ 1시간 간격으로 기후변화 조건을 고려하여 센서 지도 구축
 - 이들 데이터를 종합적으로 분석하여 실시간 교통흐름을 바탕으로 정확한 네비게이션 서비스 제공
- 미래의 발생 가능성을 예측할 수 있는 시스템 구현
 - 5분에서 15분 간격으로 수집된 데이터를 분석하여 향후 2~24시간을 예측 가능
 - ※ 밀라노 도시 전체 정보를 분석하기 위한 소요시간은 0.1초에 불과'
 - 데이터가 사전에 입력되지 않았지만, 갑작스럽게 발생하는 정보는 실시간으로 수집하여 데이터화 시켜 시스템에 반영



□ 구글의 실시간 자동번역시스템

- 구글은 기존의 방법과 달리 통계적 기법을 활용한 실시간 자동번역시스템을 개발
 - 기존의 방법과 달리 전문가가 번역한 문건을 데이터베이스화 하여 비슷한 문장과 어구를 대응시키는 통계적 기법 활용
 - ※ 기존의 일반적 번역시스템 구조: 컴퓨터에게 명사, 동사와 같은 구조와 음운을 이해시키는 방식
 - 구글은 6개국어로 번역된 유엔 회의록과 23개국어로 번역된 유럽의회 회의록을 번역 엔진에 입력
 - 또한 서적 스캐니 프로젝트(Scany Project)에서 수천만권의 전문 번역 데이터베이스를 구축



- 구글은 Spell-check 및 음성인식에도 AI기술을 적용
 - 구글은 MS가 장기간 대규모 투자를 통해 만들어진 스펠링 교정보다 우수한 프로그램을 매일 3억건씩 발생하는 '오타 입력과 수정정보'를 활용하여 개발
 - 또한 음성인식 능력향상을 위해 사용자 자율교정 정보를 페이스북에 만들고 이를 활용

IV

결론 및 시사점

- 경제·사회·기술적 환경변화의 소용돌이가 급격히 빨라지면서 미래의 다양한 가능성과 불확실성에 대한 전망이 더욱 중요)
 - 미래를 예측하여 위험에 대응하고 성장기회를 모색하는 능력이 국가의 생존과 번영을 위한 최우선 과제
 - 빅데이터는 방대한 자료의 분석을 통해 미래를 추론하고 예측하는 핵심 기반

※ 의미없어 보이는 대용량의 데이터에서 일정한 ‘패턴’을 발견하고 ‘통찰력’을 얻는 것이 빅데이터의 핵심

< 미래사회의 특성과 빅 데이터의 역할¹⁰⁾ >

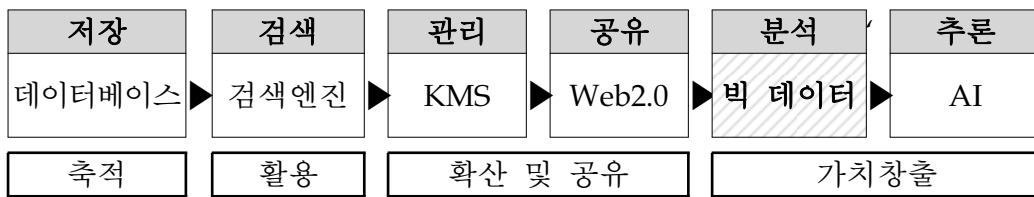
미래사회 특성		빅 데이터의 역할	
불확실성	→	통찰력	<ul style="list-style-type: none"> • 사회현상, 현실세계의 데이터를 기반으로 한 패턴분석과 미래전망 • 여러 가지 가능성에 대한 시나리오 시뮬레이션 • 다각적인 상황이 고려된 통찰력을 제시
리스크	→	대응력	<ul style="list-style-type: none"> • 환경, 소셜, 모니터링 정보의 패턴 분석을 통한 위험징후, 이상 신호 포착 • 이슈를 사전에 인지·분석하고, 빠른 의사결정과 실시간 대응 지원
스마트	→	경쟁력	<ul style="list-style-type: none"> • 대규모 데이터 분석을 통한 상황인지, 인공지능 서비스 등 가능 • 소셜(니즈)분석, 평가, 신용, 평판 분석을 통해 최적의 선택 지원 • 트렌드 변화 분석을 통한 제품 경쟁력 확보
융 합	→	창조력	<ul style="list-style-type: none"> • 타분야와의 결합을 통한 새로운 가치창출(의료 정보, 자동차정보, 건물정보, 환경정보 등) • 방대한 데이터 활용을 통한 새로운 융합시장 창출

9) 한국정보화진흥원, ‘미래사회의 새로운 가능성과 ICT의 역할’, IF보고서, 2010년 1호

10) 한국정보화진흥원, ‘성공적인 빅데이터 활용을 위한 3대 요소: 자원, 기술, 인력’, IF보고서, 2012년3호

- 빅데이터의 가능성을 극대화시키고 새로운 가치를 끌어내기 위해서는 AI기술과의 접목이 필수
 - 빅데이터의 진정한 가치는 커다란 데이터 자체에 있는 것이 아니라, 추출할 수 있는 새로운 사회적 가치가 있으며 이를 실현하는 핵심 기술이 AI
 - 기계학습, 에이전트, 자연어처리, 패턴인식 등의 AI기술은 빅데이터 처리를 위한 핵심 기술이자 기반
- ※ 구글의 CEO인 래리 페이지는 '인터넷을 거대한 인공지능으로 만드는 것이 구글의 목표'라고 밝히며, AI와 빅데이터의 결합을 활발히 추진 중

< 데이터의 과거-현재-미래 >



※ '新가치창출 엔진, 빅 데이터의 새로운 가능성과 대응 전략'(NIA, '11)보고서 재구성

- 현재 우리나라의 경우 '빅데이터 버블'이라고 불리울 정도로 빅데이터에 대한 관심은 뜨겁지만, 이를 더욱 가치 있게 하기 위한 'AI' 기술에 대한 투자는 매우 미흡한 실정
 - 데이터의 단순한 활용을 넘어, 빅데이터를 제대로 이해하고 활용하는 기업은 소수에 불과
 - ※ 빅데이터를 보통이상 알고 있는 경영자 비율은 56.4%이나 적극 활용하는 경우는 19.4%(삼성경제연구소, 2012)
 - AI분야는 고착상태로 해외 의존도가 높아 선진국에 비해 경쟁력이 낮고, 최근 산업 침체로 인한 업계 활력도 낮은 상황
 - AI 관련 포괄적 연구그룹도 전무하고, 정부의 정책적 지원도 미흡
- 빅데이터와 AI분야에 대한 전략적인 접근이 절시

1 빅데이터와 AI를 결합한 다양한 비즈니스 모델 개발

- 'AI'는 맞춤형 서비스를 제공하기 위한 기반 기술로, 빅데이터와의 만남으로 그 가치를 배가시킬 수 있음
 - 이미 구글, 애플 등 선진기업은 AI와 빅데이터를 접목한 다양한 서비스들을 제공 중
 - 의료, 교육, 재난·재해, 복지 등 다양한 분야의 비즈니스 모델 개발 및 시범사업 추진이 필요

2 AI기술 및 인력의 전략적 육성

- 최근 이공계 기피현상과 함께 컴퓨터공학 전반에 대한 관심이 저조한 상황
 - 특히 AI분야는 가능성에 대한 비관적인 관점의 확대 등으로 인해 국내 기반 자체가 매우 미흡한 실정
 - 빅데이터를 필두로 진행되고 있는 'Smart 물결'에 능동적으로 대응하기 위해서는 AI기술 및 인력에 대한 전략적 육성이 필수
- ※ 국내 국책 연구소에 인공지능 분야를 담당할 수 있는 부서의 신설 등의 통해 체계적인 AI 연구 추진

3 기술 + 사회문화적 요소에 대한 면밀한 검토

- 빅데이터 및 AI를 논의에서 빠지지 않는 것이 개인정보 유출 및 사생활 침해와 같은 역기능 문제
 - 따라서 기술적 측면으로만 접근하면 실질적인 파급효과 및 실현 가능성은 저하될 수밖에 없음
- 기술의 발전과 더불어 우리가 원하는 미래상에 대한 명확한 이해와, 이를 달성하기 위한 정치사회적 기반에 대한 근본적 모색이 중요



참고 자료

- [1] 국가정보화전략위원회(2011), '빅 데이터를 활용한 스마트 정부 구현'
- [2] 노무라연구소(2012), '빅데이터 시대 도래', IT 프론티어 3월호
- [3] 데이터넷(2011), '급부상하는 '비즈니스분석', 더 넓고 깊은 통찰력 필요'
- [4] 문혜정(2012), 'Big Data 구축기술과 사례를 중심으로'
- [5] 오성훈(2012), '빅데이터 기술 현황', 한국저작권위원회 산업연구팀
- [6] 정지선(2011), '新가치창출 엔진, 빅 데이터의 새로운 가능성과 대응 전략'
- [7] 최규현(2012) '빅데이터 연구 동향과 시사점', 정보통신산업연구원 IT기획시리즈
- [8] 고태영 (2012), '데이터 분석가가 바라본 빅데이터 현상과 대응을 위한 준비방법'
- [9] 이지영(2012), '인공지능', 빅데이터의 미래인가,
- [10] 한국정보화진흥원(2012), 빅데이터로 진화하는 세상, '빅데이터 글로벌 선진사례'
- [11] 정지선(2012), '미래사회의 새로운 가능성과 ICT의 역할'
- [12] 정지선(2012), '성공적인 빅데이터 활용을 위한 3대 요소: 자원, 기술, 인력'
- [13] Gartner(2011), 'CEO Advisory : 'Big Data' Equals Big Opportunity'
- [14] Gartner(2011), 'How to Plan, Participate and Prosper in the Data Economy'
- [15] IDG Korea(2012), '빅 데이터를 위한 개방형 DB 프레임워크 "하둡"의 이해'
- [16] IDG Korea(2012), '빅 데이터를 클라우드에서 ', 2012.3.22
- [17] IDG Korea(2011), <http://www.itworld.co.kr/news/72492>
- [18] InforWorld(2011), 'The 6 hottest new jobs in IT'
- [19] McKinsey Global Institute(2011), 'Big Data : The next frontier for innovation, competition, and productivity'
- [20] MT Slon Management Review(2011), 'Big Data, Analytics and the Path From Insights to Value'
- [21] Jeffrey Pfeffer(2006), 'Hard Facts, Dangerous Half-Truths, and Total Nonsense'

2012년도 IT & Future Strategy 보고서

- 제1호(2012. 3), 「실리콘벨리가 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), 「빅데이터시대 : 애군시스템을 enffjTKs 시장경쟁과 전략분석」
- 제5호(2012. 5), 「빅데이터 시대의 데이터 자원 확보와 품질관리 방안」
- 제6호(2012. 6), 「빅데이터시대 : 공공부문 빅데이터 추진방향」
- 제7호(2012. 7), 「빅데이터 시대, AI의 새로운 의미와 가치」

1. 본 보고서는 정보통신진흥기금으로 수행한 정보통신연구개발사업의 연구결과입니다.
2. 본 보고서의 내용을 발표할 때는 반드시 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.