

BIG DATA

INDEX

BIG DATA

1. 빅 데이터란 ? ----- p
2. 빅 데이터와 네트워크 ----- p
3. 빅 데이터의 활용사례 ----- p
4. 빅 데이터의 전망 ----- p
5. 빅 데이터의 한계 ----- p

1

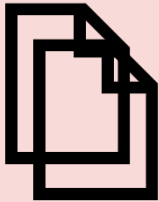
빅데이터 출현 배경

BIG
DATA

비정형 : 음성/영상 , SNS



반정형 : 로그 데이터, 머신 데이터



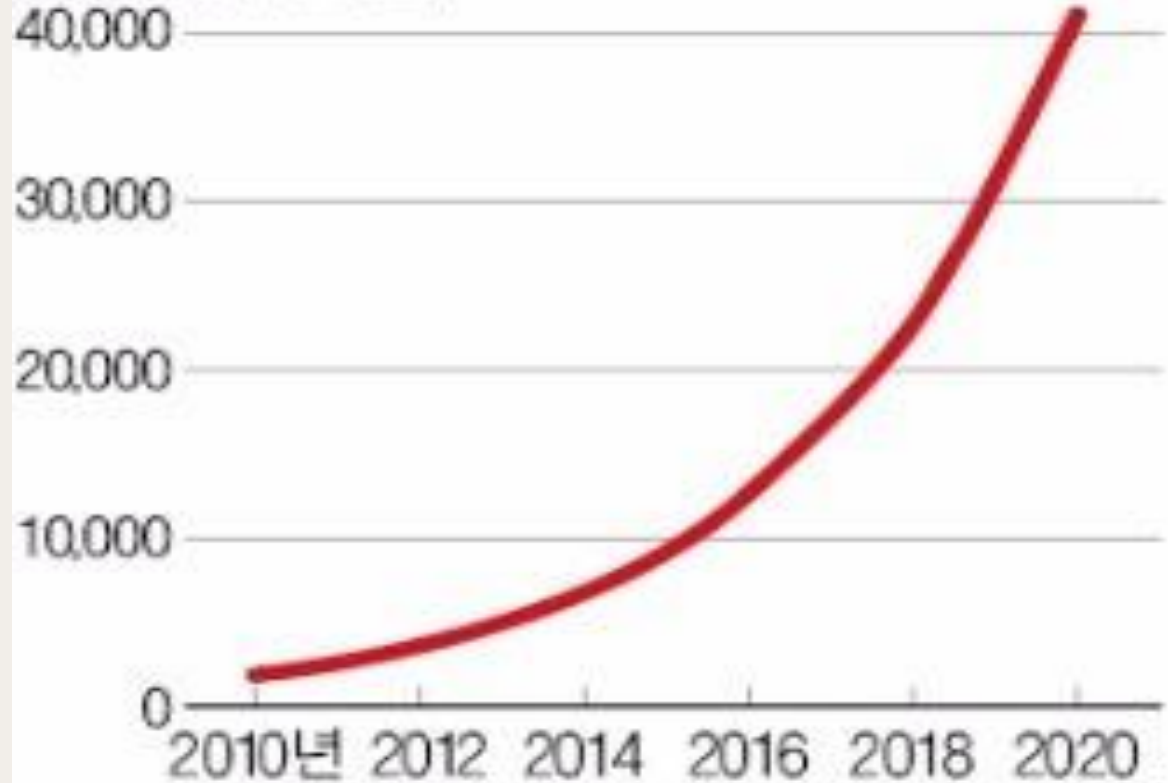
정형 : DB, DW



사진 1

급증하는 데이터

단위:엑사바이트(EB)



※1EB=104만8576테라바이트(TB)

2016년 이후는 전망

자료:EMC, IDC

1

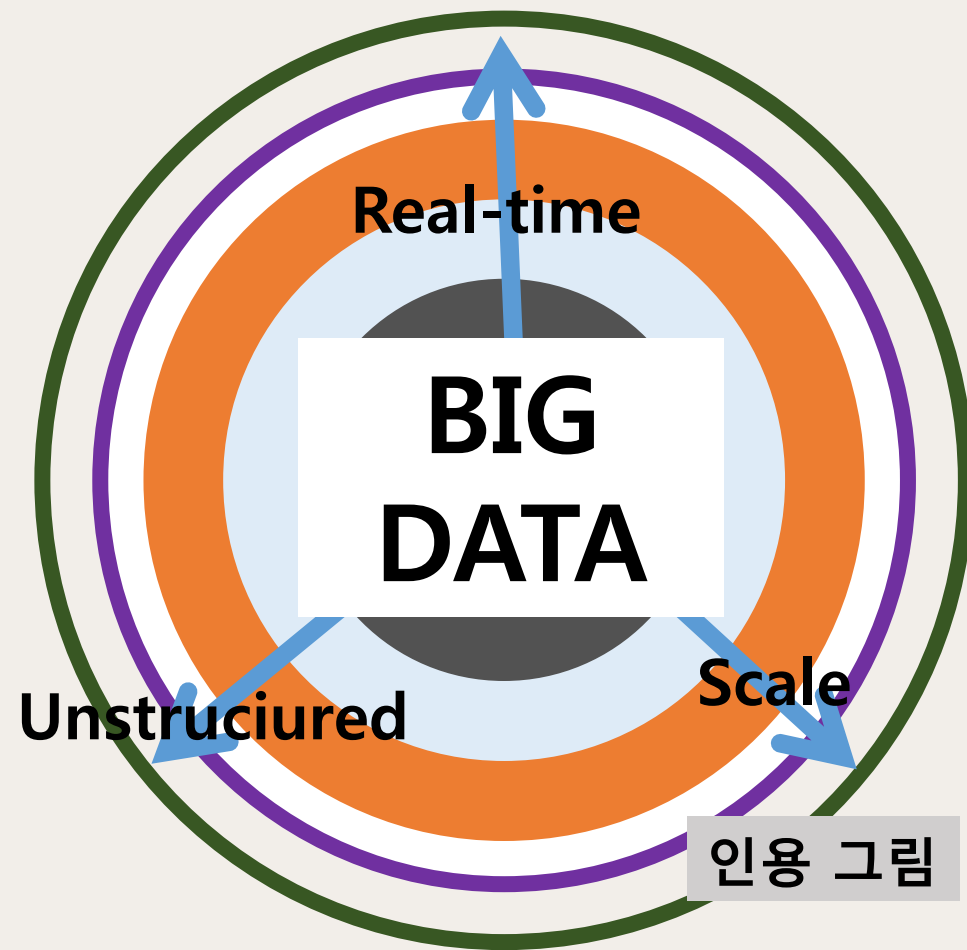
빅데이터란?

BIG DATA

플랫폼의 변화와 관련한 하드웨어
처리용량/비용상의 이점

대용량 메모리 모델로의 전환→
기회요인 활용

을 ! 위한 ! 일련의 현 상



1

빅데이터 속성

BIG
DATA

기존의 3V

Velocity – 속도 **V**ariety – 다양성 **V**olume – 양

새롭게 제시되는 4V

Veracity – 정확성 **V**ariability – 가변성 **V**isualization – 시각화

1

빅데이터의 중요성 (필요한 이유)

BIG
DATA



경쟁력 강화

적절성의 변화
서비스 기대치 예측 반영
새로운 비즈니스 모델 탄생
경쟁력 강화

사진 2

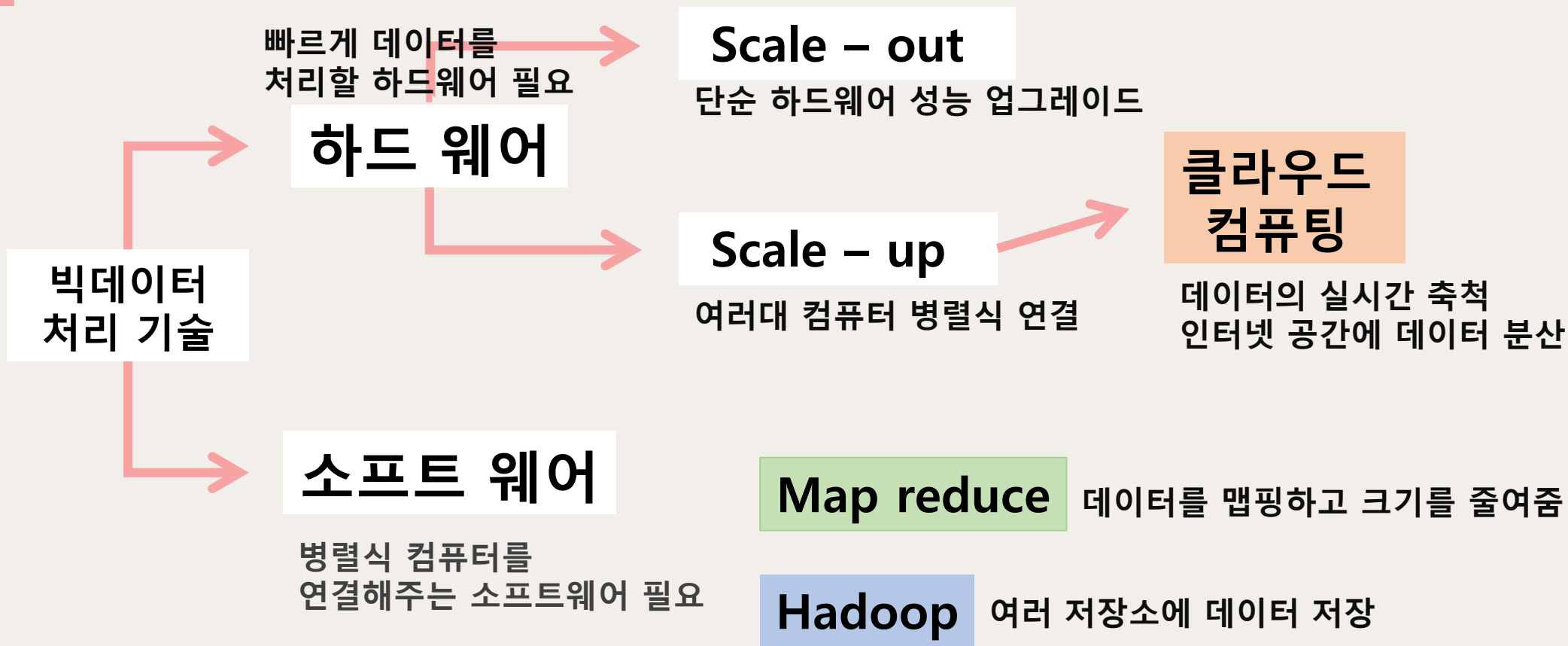


사진 3



사진 4





3

활용사례

BIG
DATA

바이오, 소셜, 생산, 금융, 통신 분야에 활용

국내

사진 5



사진 7



사진 9



국외

사진 6



사진 8



사진 10



2013년 Goodreads (회원수 2천5백만 SNS) 회사 인수 →
회원들의 SNS상에서 책에 대한 토의 논의 내용 분석 →
대다수가 좋아할만한 것을 추천

머신 러닝
기술

빅 데이터를 사용하여 가격을 최적화

경쟁 회사 가격

재고량

선호도

과거 주문 사례

빅 데이터를 분석하여

10분 간격 업데이트

매년 25%
이익 상승

3

활용사례

BIG
DATA

Google

플루 트렌드(FLU TRENDS)

→ 목적 : 이를 통해 독감 발병을 예측

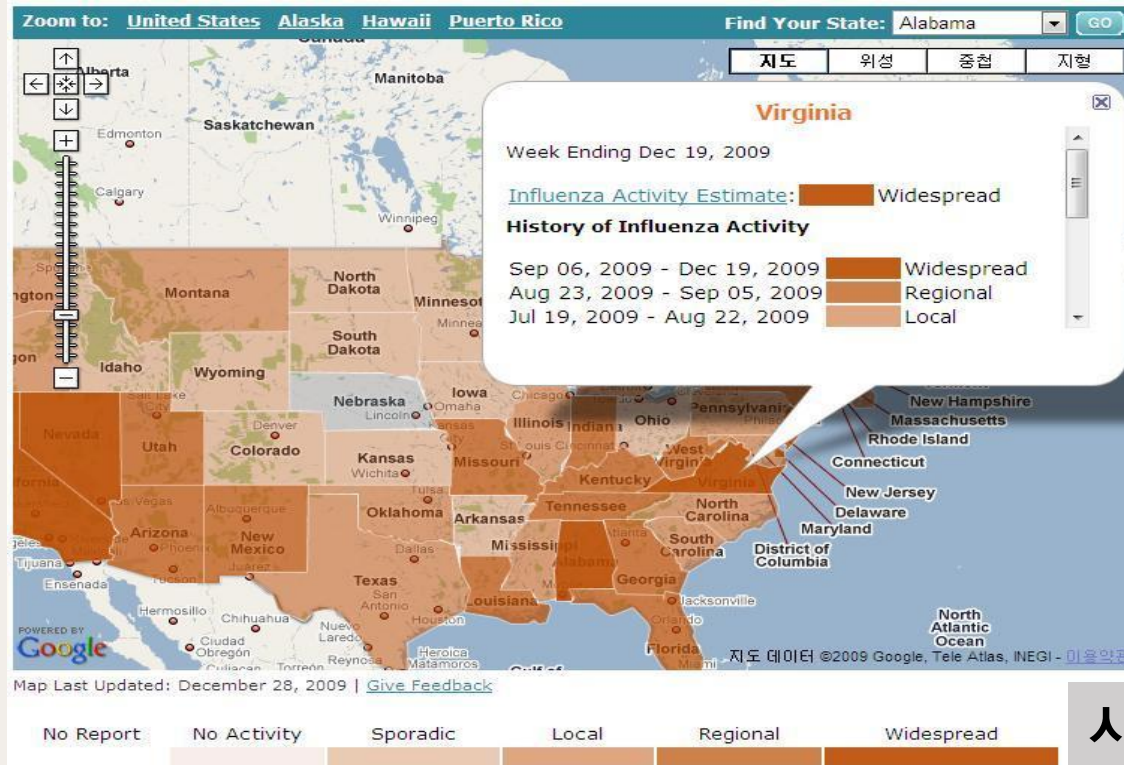


사진 11

ZARA



사진 12



*“ 다양한 빅데이터 기반의 ”
서비스 모델을 새롭게 개발*

고객

SNS

소셜 데이터

검색 키워드



고객의 사회 트렌드 파악

기업의 이미지 파악

연예인 활동 및 이미지 분석

모바일 광고 분석

소셜데이터 분석



방대한 고객의 행동 및 로그 데이터를 분석

“사지 탐지(Fraud detection)” 알고리즘

* 병원 의료데이터 활용

건강관리 제안

치료방법 제안

질병 예측

치료중심이 아닌 관리 및 예방..

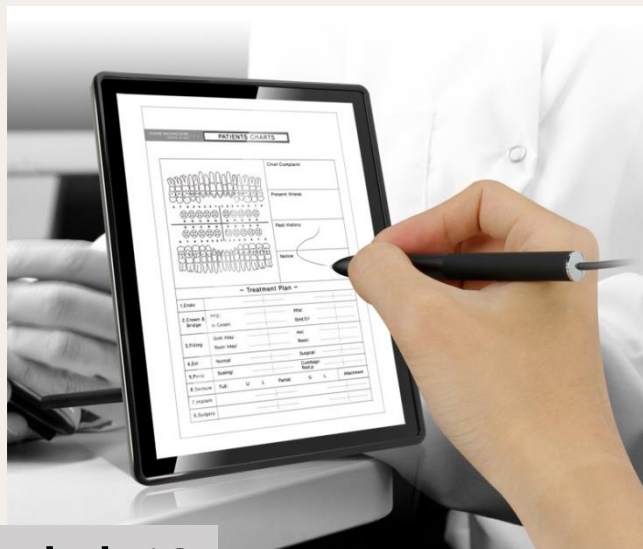


사진 13

< 의료 차트 >

불법행위
모니터링

일련번호	환자명	성별	생년월일	진단명	진단일자	처방명	처방일자	처방량	처방기간	위치	입원진단	입원일자	입원종료일자
1	1543	남	1980	조	2000-07-21	사	2000-07-20	10:15	15시간	C-14	AC of 2000 or r/o RFE, A-11b	2000-07-20	20:55
2	0040	남	1981	조	2000-07-20	남	2000-07-20	10:21	20시간	D-16	DM(+)-Diabetic colitis active	2000-07-20	20:55
3	1164	남	1986	조	2000-07-22	남	2000-07-21	10:29	20시간	D-22	r/o Hematoma, M, s/p Cerebral	2000-07-21	10:36
4	2091	남	1983	조	2000-07-22	남	2000-07-21	11:42	20시간	환상2-14	Hyponatremia, NPO, CLO, TND) s/p CCR	2000-07-21	10:36
5	0061	남	1986	조	2000-07-21	남	2000-07-21	10:00	20시간	환상2-11	Hyponatremia, NPO, CLO, TND) s/p CCR	2000-07-21	10:36
6	2115	남	1980	조	2000-07-22	남	2000-07-21	11:26	20시간	A-85	Hyponatremia with ureteric colic	2000-07-21	20:00
7	1974	남	1981	조	2000-07-22	남	2000-07-21	11:22	20시간	D-10	hematochezia/chronic DVI s/p	2000-07-21	17:12
8	2170	남	1986	조	2000-07-21	남	2000-07-21	11:14	20시간	환상2-21	AML	2000-07-21	20:15
9	2179	남	1980	조	2000-07-22	남	2000-07-21	11:53	20시간	A-84	Renal tuberculosis	2000-07-21	19:24
10	2152	남	1980	조	2000-07-24	남	2000-07-21	11:19	20시간	A-85	IP site oozing/po[13/3/6] 118	2000-07-21	12:00
11	2179	남	1980	조	2000-07-24	남	2000-07-21	11:21	20시간	환상2-01	Hyponatremia, NPO, CLO, TND) s/p CCR	2000-07-21	10:36
12	0372	남	1986	조	2000-07-23	남	2000-07-22	10:45	20시간	C-10	W/O/OO on CAP	2000-07-22	10:36
13	1953	남	1986	조	2000-07-22	남	2000-07-22	10:45	20시간	D-19	W/O/OO on CAP	2000-07-22	10:36
14	2143	남	1986	조	2000-07-22	남	2000-07-22	11:00	20시간	환상2-16	Hyponatremia, NPO, CLO, TND) s/p CCR	2000-07-22	10:36
15	2179	남	1980	조	2000-07-22	남	2000-07-22	11:12	20시간	환상2-08	Anterior mediastinal mass	2000-07-22	10:36
16	1523	남	1986	조	2000-07-22	남	2000-07-22	11:27	20시간	환상2-15	Seizure, MCC s/p brain, sag	2000-07-22	10:36
17	2142	남	1986	조	2000-07-22	남	2000-07-22	11:46	20시간	환상2-11	Hyponatremia, NPO, CLO, TND) s/p CCR	2000-07-22	10:36
18	2104	남	1987	조	2000-07-23	남	2000-07-22	11:54	20시간	환상2-01	Hyponatremia, NPO, CLO, TND) s/p CCR	2000-07-22	10:36
19	0614	남	1986	조	2000-07-22	남	2000-07-22	11:06	20시간	환상2-10	Hyponatremia, NPO, CLO, TND) s/p CCR	2000-07-22	10:36
20	2170	남	1980	조	2000-07-22	남	2000-07-22	11:13	20시간	C-09	Hyponatremia, NPO, CLO, TND) s/p CCR	2000-07-22	10:36
21	1905	남	1986	조	2000-07-22	남	2000-07-22	11:47	20시간	환상2-15	Hyponatremia, NPO, CLO, TND) s/p CCR	2000-07-22	10:36

사진 14

< 간호 기록 >

3

활용 사례

BIG DATA

(1테라바이트 기업용 2000만원) X 2048
= 4096000만원

2페타바이트

제임스 캐머런

15년 전의 스토리 → 당시 시대 구현 불가
엄청난 규모의 영상제작에 들어가는 **데이터**

이를 수용하는 **스토리지**

3D화 작업의 데이터 처리 **컴퓨팅 파워**

SF 특수 효과 처리 → 여러 차례 **렌더링**



클라우드 컴퓨팅 도입

미디어 업종에서 클라우드와 빅데이터 성 플랫폼 적용한 첫 사례



사진 15



사진 16

빅데이터 + 다양한 기술 => 무궁무진한 **가능성!!**

이런 전망에 빅데이터의 시장은 더욱 **HIGH**



2011~2026년간 빅데이터 시장 규모 전망 (단위: 10억 달러)

<출처: statista>



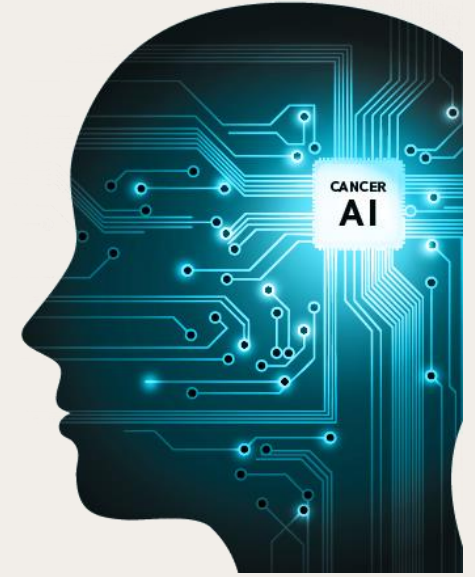


사진 18

<페이팔의 딥러닝 사용>



학습

컴퓨터
인공지능망

컴퓨터의 **인공지능망**과 **빅데이터**가 융합하여
컴퓨터가 스스로 **학습**하는 것

다양하고 방대한
DATA 일수록 딥러닝 성능이 중요

알고리즘의 진화

알고리즘 : 문제 해결을 위한 방법

사용자가 입력한 검색어에서 가장 적합한 결과 제공이나 가장 싸거나, 최신의 결과를 알려주는 수준의 알고리즘 제공.

<Before>

빅데이터

빅데이터를 통해 사용자 개개인에게 맞춤형 서비스와 같은 고도의 진화된 알고리즘 제공 가능.

<After>

웨어러블 디바이스와 IOT

사진 18

웨어러블 디바이스
(착용 컴퓨터)

사진 18

IOT
(사물 인터넷)

사진 18



기존에 데이터로 파악하지 못한

인간의 감정까지 측정이 가능

사용자의 상태에 따라 사물인터넷의 반응

빅 데이터는 만능일까?

엄청난 가능성을 지닌 기술
BUT, **만능의 열쇠 X**

빅 데이터의 분석

빅데이터는 그 **자체만으로는 가치 X**



빅데이터를 **분석**하고 **활용**하는 것은 결국 사람.
빅데이터 자체도 중요하지만 **분석 기술** 또한 중요

독감 트렌드 - 미국

독감 유행 수준을 파악할 때 검색어가 좋은 지표가 될 수 있습니다. Google 독감 트렌드는 집계된 Google 검색 데이터를 사용하여 독감 유행 수준을 예측합니다. [자세히 알아보기 >](#)

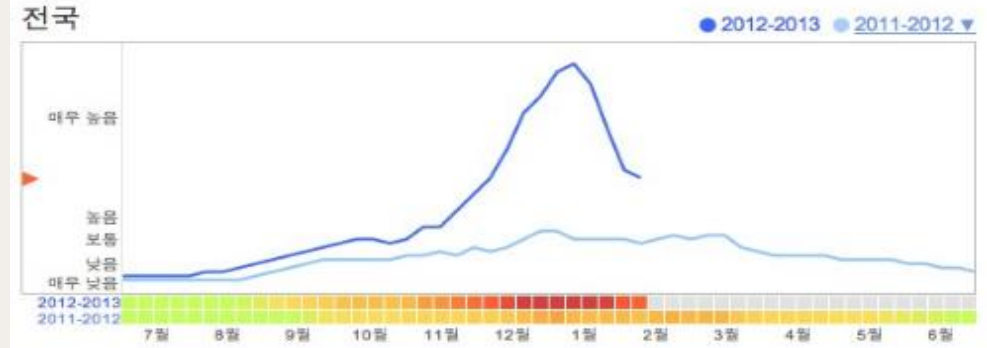


사진 19

<구글의 플루 트렌드 프로젝트>



사진 20

<영화 마이너리티 리포트>

빅데이터와 딥데이터

빅데이터라 해서 무조건 가치있는 정보일까?

⇒ 양만 많은 **쓰레기** 데이터들이거나 한 쪽에 **편중된** 데이터면 원하는 결과를 도출하기 힘들 수 도 있다.

딥데이터란? : 대용량 데이터를 기반으로 하는 것은 빅데이터와 유사하지만 빅데이터는 사소한 정보 하나하나를 다 수집하는 반면 딥데이터는 조사하고자 하는 것을 기준으로 필요한 것만 **집중적**으로 수집.

⇒ 필요한 데이터만 뽑아주므로 결론 도출에 시간과 비용을 절약할 수 있음.

⇒ 빅데이터 + 딥데이터 둘 다 시기적절하게 잘 사용하면 훨씬 좋음!!

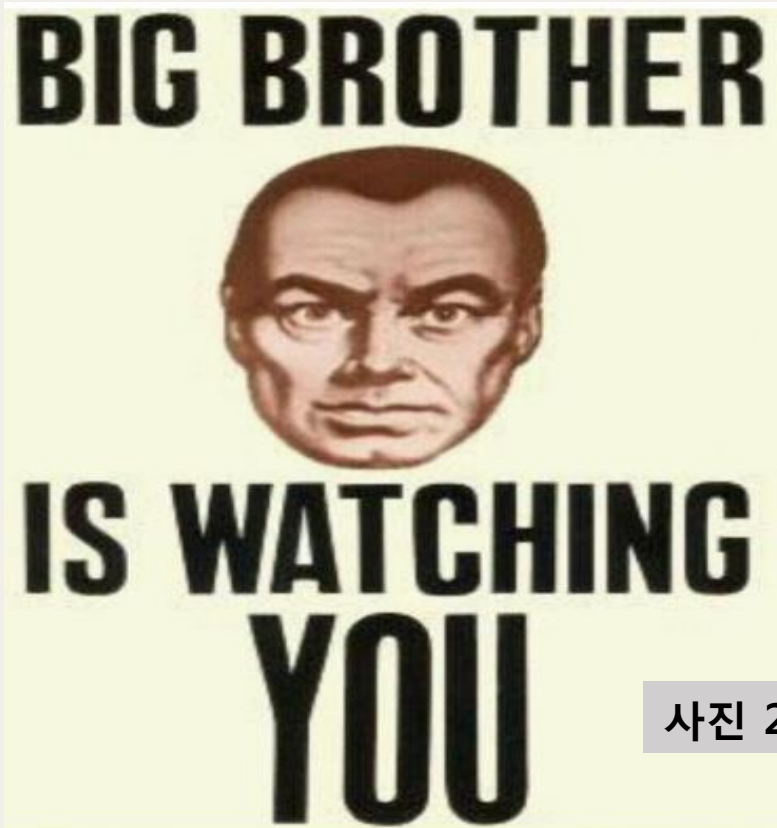


사진 21

조지오웰 '1984'의 빅브라더

중앙기관에 의한 완전한 통제 경고

개인의 사생활 침해 가능성
악용 시 개인, 크게는 사회에 영향

잊고 싶은 기억이 데이터로 영원히 저장?



데이터 감옥

→ 데이터의 유효기간과
같은 제도적 방안도 필요

제도에 막혀 기술 발전의 저해도 사실 → 허용과 제한의 적절한 제도적 **균형** 필요!!

윤리적 교육 강화 필수 => 빅데이터가 개인에게 미칠 수 있는 영향력을 생각하면 기술보다 윤리적인 면이 더욱 필요.

빅데이터 만능주의 탈피 => 빅데이터가 더 좋을 때가 있고 과거의 기술이 더 좋을 때가 있다.

빅데이터는 결국 도구!! => 빅데이터는 결국 도구이다. 그것을 사용하는 인간의 의지와 방향에 따라 빅데이터는 독이 될수도 약이 될수도 있을 것이다.

출 처

참고문서 : 도서명 : 빅데이터@워크/ 저자 : Davenport, Thomas H. / 출판사 : 21세기 북스

도서명 : 빅데이터 승리의 과학 / 저자 : 고한석 / 출판사 : 이지스퍼블리싱 /

도서명 : 빅데이터 분석 개론 / 저자 : 안동혁, 김선영 / 출판사 : 보리별

도서명 : 빅데이터 세상 / 저자 : 매일경제 기획팀, 서울대 빅데이터 센터/ 출판사 : 매일 경제 신문사

도서명 : 빅데이터를 처형하라 / 저자 : 배철순 / 출판사 : 프리덤 월드

빅데이터 개념 : <http://www.bigdata.go.kr/intro.html>

빅데이터 출현 배경 : <http://subinne.tistory.com/38>

빅데이터 기술 : <http://blog.lgcns.com/1403#recentEntries>

빅데이터 정의 : <http://www.bigdata.go.kr/>

빅데이터 요소 : <https://terms.naver.com/entry.nhn?docId=3386305&cid=58370&categoryId=58370>

빅데이터 중요성 : <https://www.sktinsight.com/98202>

빅데이터 개요 : <http://goodaum.co.kr/221222802925>

클라우드 컴퓨팅 : https://blog.naver.com/suhyup_bank/221186832740

클라우드 : <http://gotocloud.co.kr/?p=564>

딥러닝 : blogs.nvidia.co.kr/2016/08/03/difference_ai_learning_machinelearning

웨어러블 디바이스 : https://ko.wikipedia.org/wiki/%EC%B0%A9%EC%9A%A9_%EC%BB%B4%ED%93%A8%ED%84%B0

IOT : <https://ko.wikipedia.org/wiki/사물인터넷>

딥데이터 : <https://translate.google.co.kr/translate?hl=ko&sl=en&u=https://channels.theinnovationenterprise.com/articles/the-difference-between-big-data-and-deep-data&prev=search>

사진 출처

사진 1 : 급증하는 데이터 - <http://goodaum.co.kr/221222802925>

사진 2: <http://www.yonhapnews.co.kr/bulletin/2018/04/23/0200000000AKR20180423046400004.HTML?input=1195m>

사진 3 : <http://www.datanet.co.kr/news/articleView.html?idxno=122520>

사진 4 : <http://www.kyeonggi.com/?mod=news&act=articleView&idxno=1459077>

사진 5 : SK: <http://www.sktelecom.com/view/advertise/brand.do>

사진 6 : 아마존 사진 출처: <http://it.donga.com/27544/>

사진 7 : 국민카드: <http://www.hansup.or.kr/home>

사진 8 : 자라 사진 출처: <http://www.naeilmohaji.co.kr/news/articleView.html?idxno=273>

사진 9 : NC 소프트 <http://game.donga.com/77387/>

사진 10 : 구글 사진 출처 <https://pixabay.com/ko>

사진 11 : 플루 트렌드 <http://cnrcblog.tistory.com/2>

사진 12 : 자라 사진 : <https://www.forbes.com/forbes/welcome/?toURL=https://www.forbes.com/companies/zara>

사진 13 : 의료 차트 : <http://www.itworld.co.kr/print/77298>

사진 14 : 간호 기록 : http://www.samsunghospital.com/dept/main/index.do?DP_CODE=ER&MENU_ID=002007

사진 15 : 아바타 <http://news.joins.com/article/21345002>

사진 16 : <https://cbmpress.com/vancouver>

사진 17 : 빅데이터 전망 - <http://www.elec4.co.kr/article/articleView.asp?idx=15220>

사진 18 :페이팔- www.elec4.co.kr/article/articleView.asp?idx=15220

사진 19 : 구글 플루트렌드 – <http://www.bloter.net/archives/143856>

사진 20 : 마이너리티 리포트 - <http://www.hani.co.kr/arti/international/america/688330.html>

사진 21 : 빅브라더 <https://m.blog.naver.com/PostView.nhn?blogId=dinglee&logNo=220638519511&proxyReferer=https%3A%2F%2Fwww.google.co.kr%2>