

# 사회물리학 빅데이터를 중심으로

물리학과 첨단기술의 세계 2011년 5월 제20권 5호 정우성

〈첨단기술과 진로탐색〉

물리학과 이시연

2015550024

# 목차

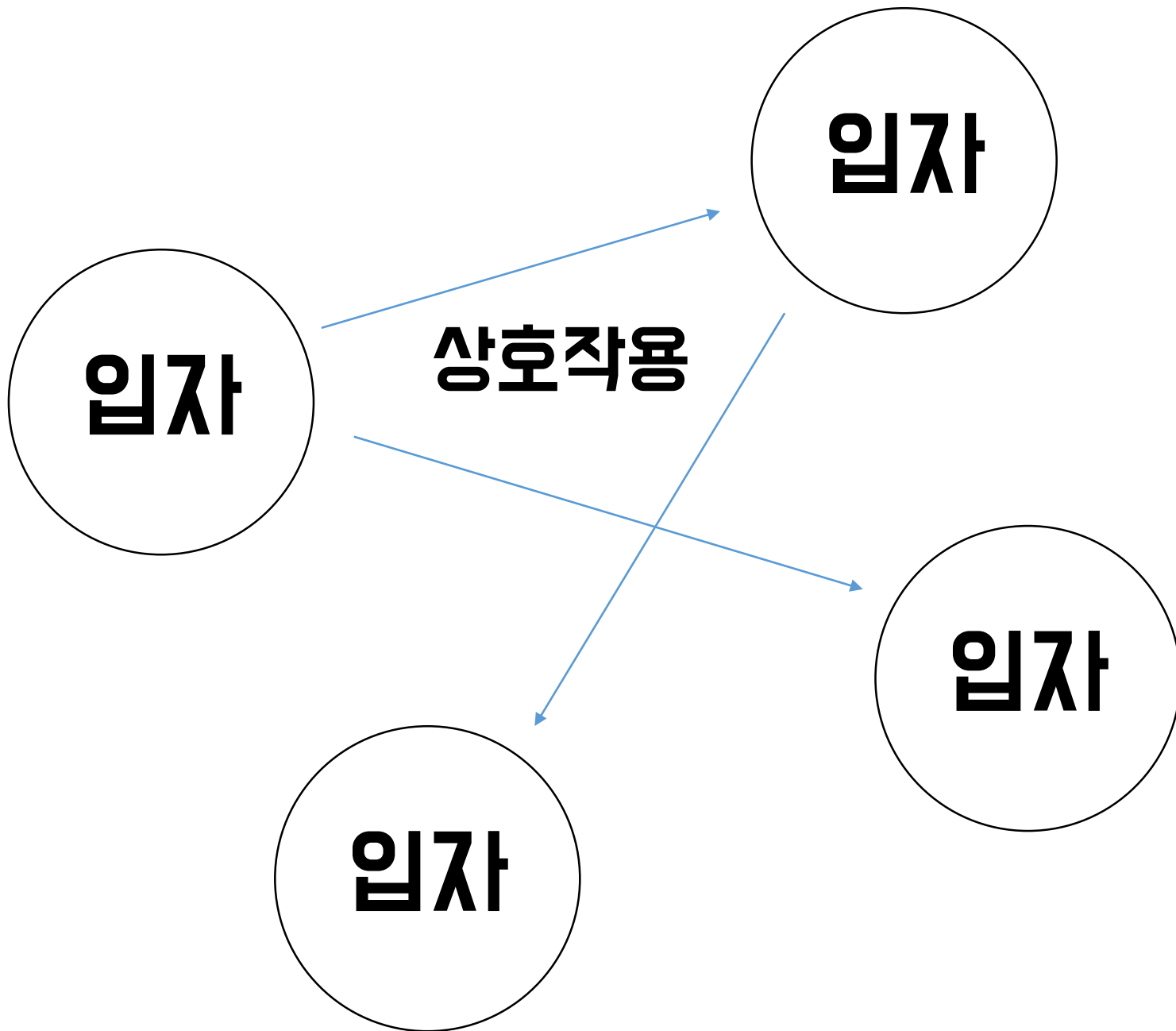
1. 주제 소개
2. 여러가지 개념 - 빅데이터와 구글 GET
3. 관련 기술 - 리버스 엔지니어링
4. 참고 문헌 및 출처

# 사회물리학이란?

주제 소개

계

보편성?



# 주식시장

보편성?



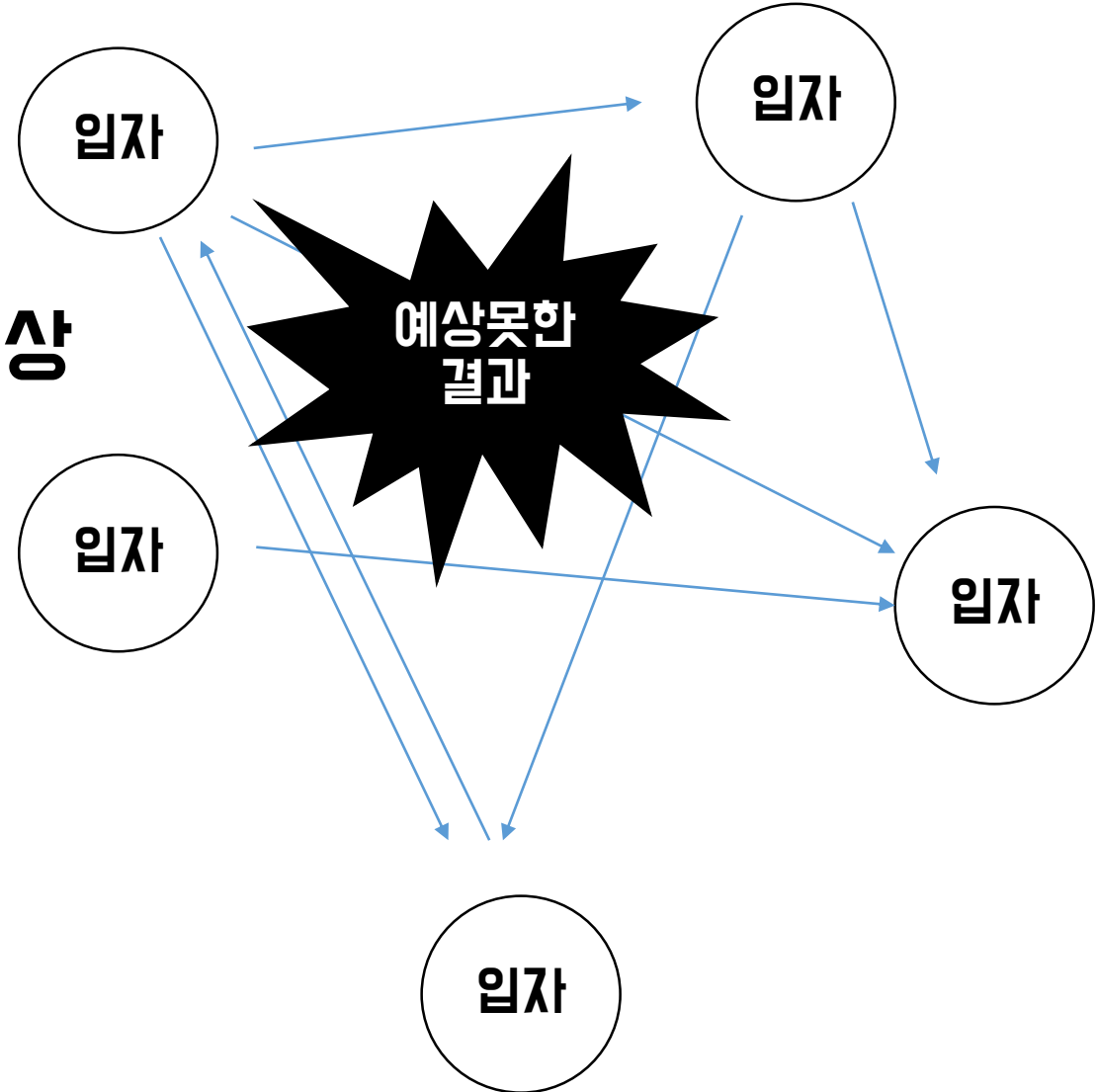
정보제공



**여러 가지 개념**

# 창발효과

- 전체는 부분보다 크다
- 입자간의 상호작용으로  
예상 못한 결과가 발생하는 현상



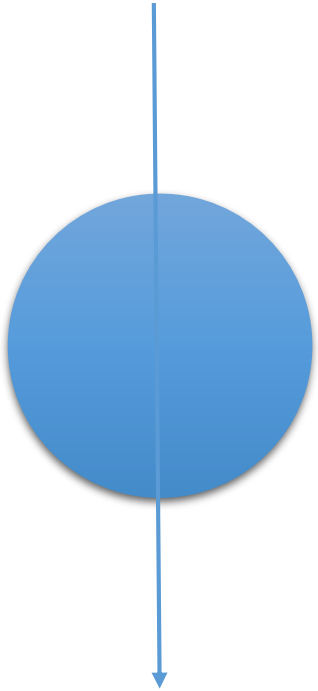
# 게임 이론



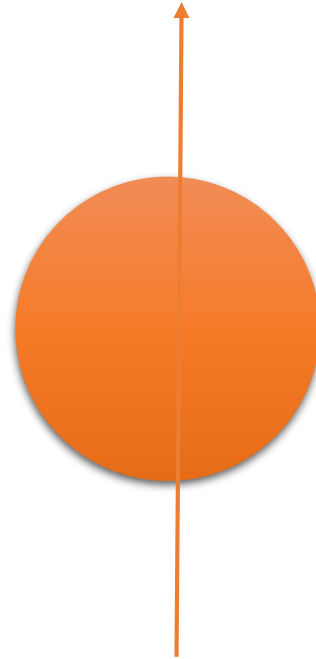
1. 규칙 아래서 진행
2. 정보가 중요
3. 전략, 전술이 핵심
4. 결과에 따른 보수 존재



# 이징 모형

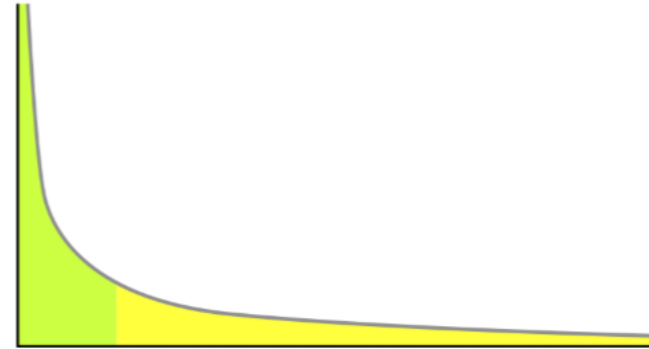
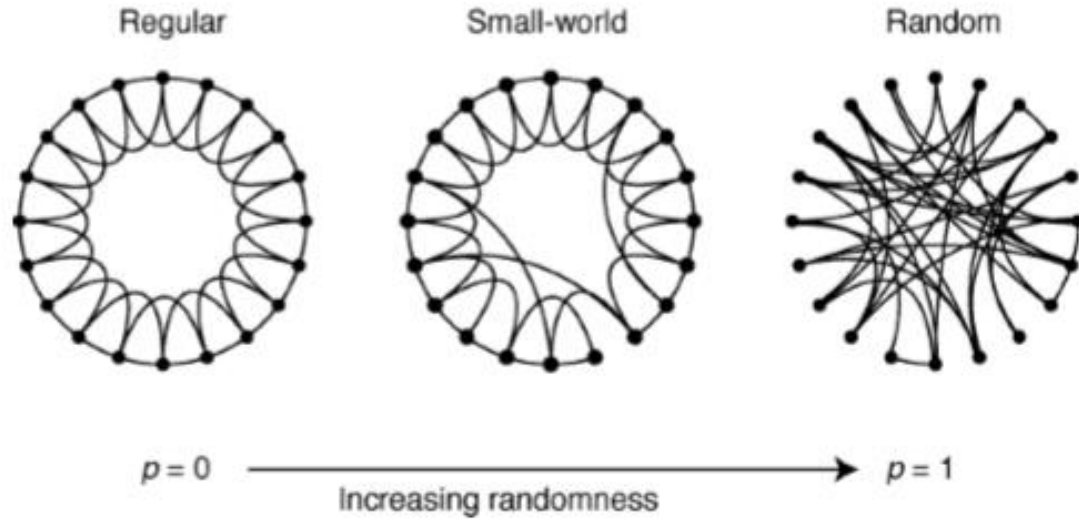


**down spin**  
**Ex. sell**



**up spin**  
**Ex. buy**

# 복잡계 연결망



작은 세상 모형

마당발  
- 거둢 제공 분포

# 빅데이터 기업의 실제 이용 사례

구글 GET의 독감 확산 분석을 중심으로

# 빅데이터의 특징

- 인과관계 → 상관관계

- 데이터의 양 뿐 아니라 데이터의 질 변환

- 지금까지 수집 불가, 대상 미포함, 분석 수단 없음으로 버려지던 모든 자료 포함

- 생활의 모든 측면에서 “데이터 배기가스” 생성!!

- 메시지 교환, 상품 주문, 파일 공유 등등.....

- 자료 규모(MORE)    자료구성(MESSY)    분석준거 - 근본적 변화

# 빅데이터의 3대 변화

- 자료 규모

- 기존의 표집에서 관찰 불가능한 소수 하위집단 세분화 관찰 가능

- 자료 구성

- 엄밀한, 통제된 정확성을 요구, 충족 불가능

- 분석 준거

- 인과관계 → 상관관계



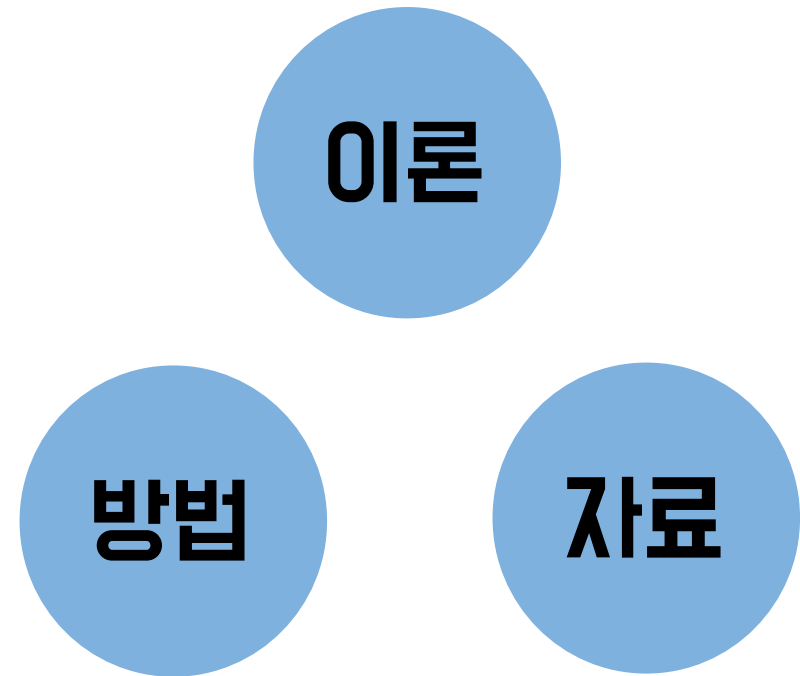
**“데이터의 홍수로 인해  
전통적인 과학적 연구방법은  
쓸모없게 되었다.”**

**크리스 앤더슨, 이론의 종말, <와이어드>, 2008**

# 학계의 인식 - 결론

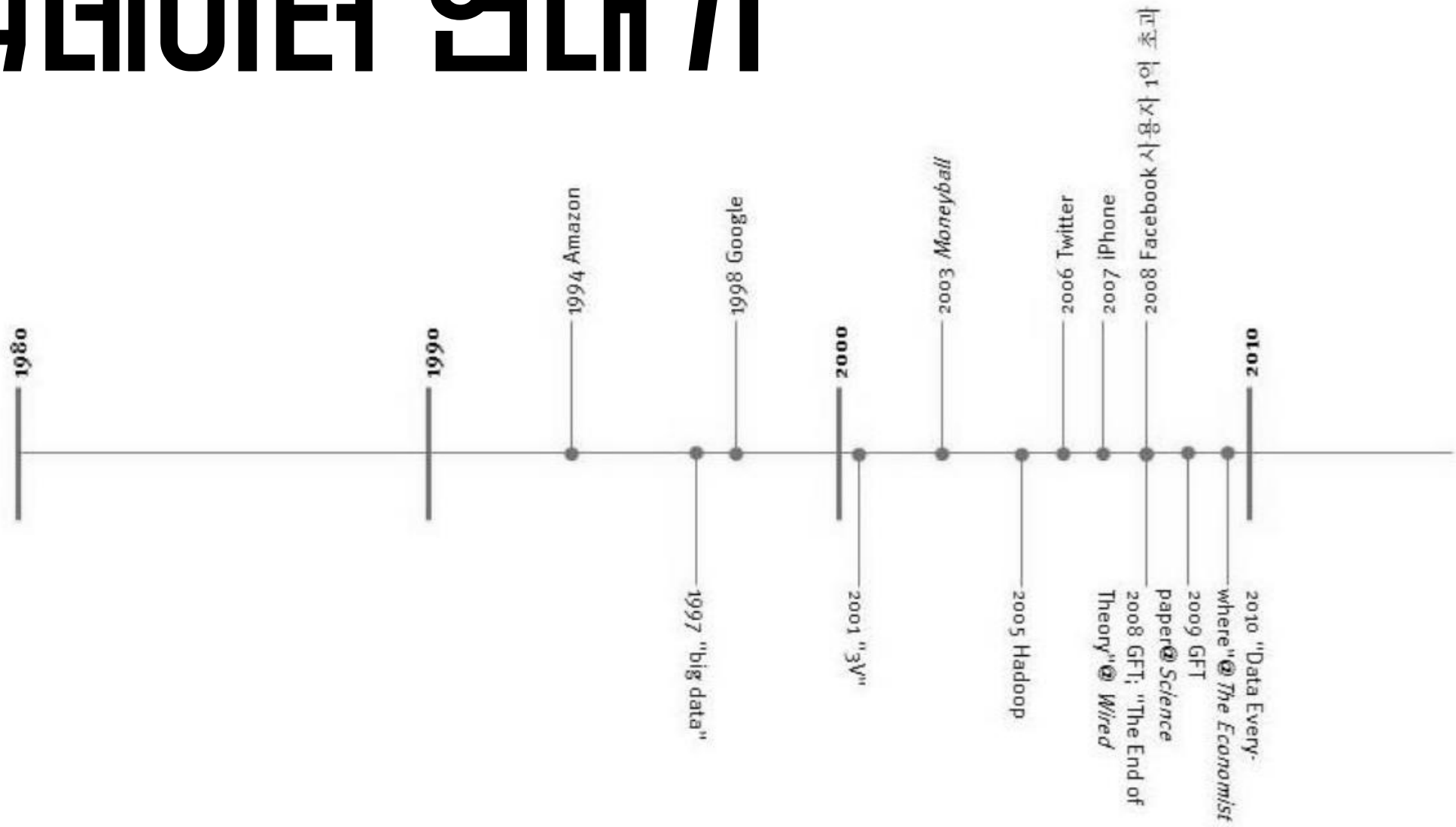


**단순한 많은 데이터 아님!!!**



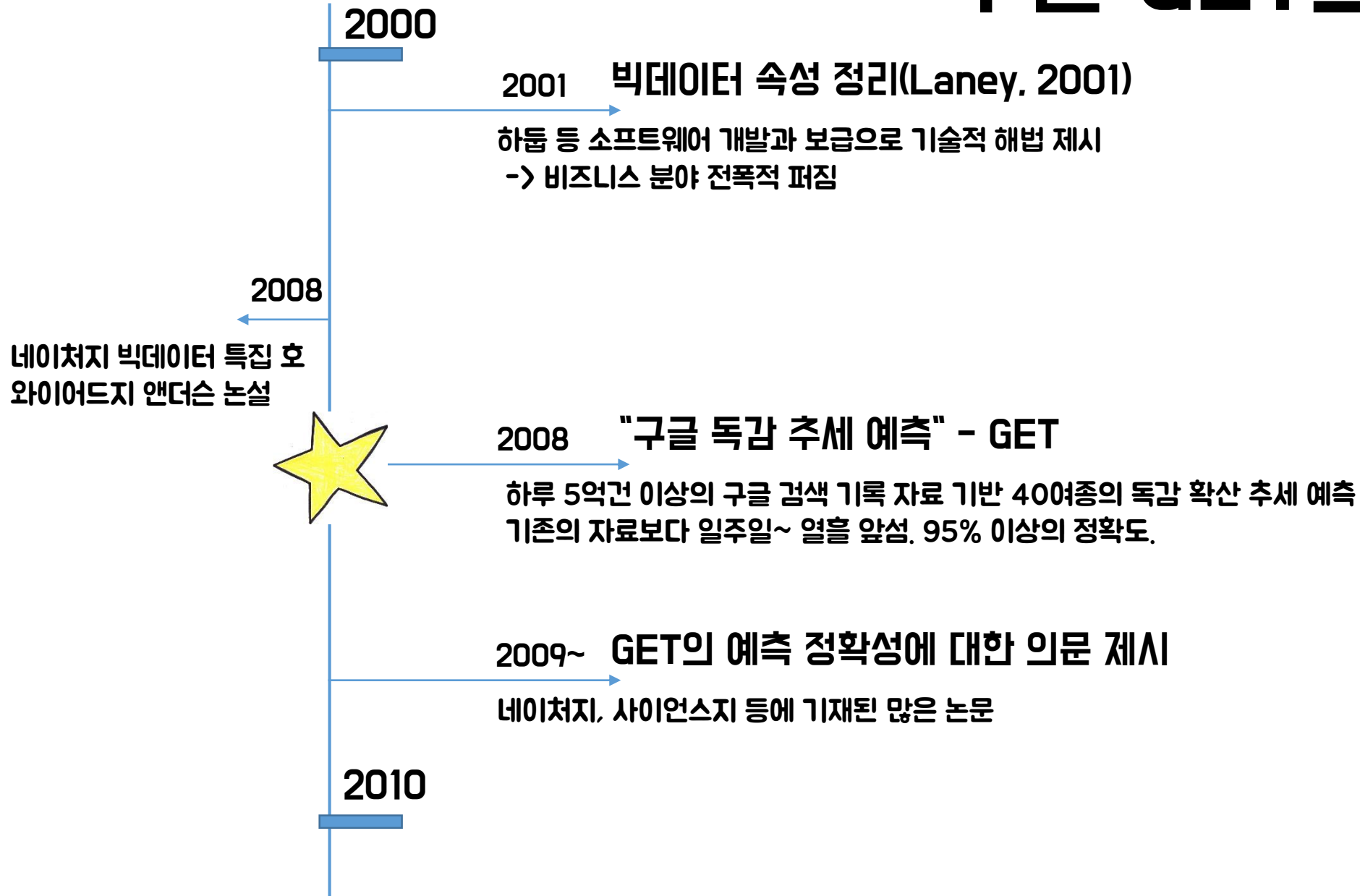
**복합적인 상호작용**

# 빅데이터 연대기

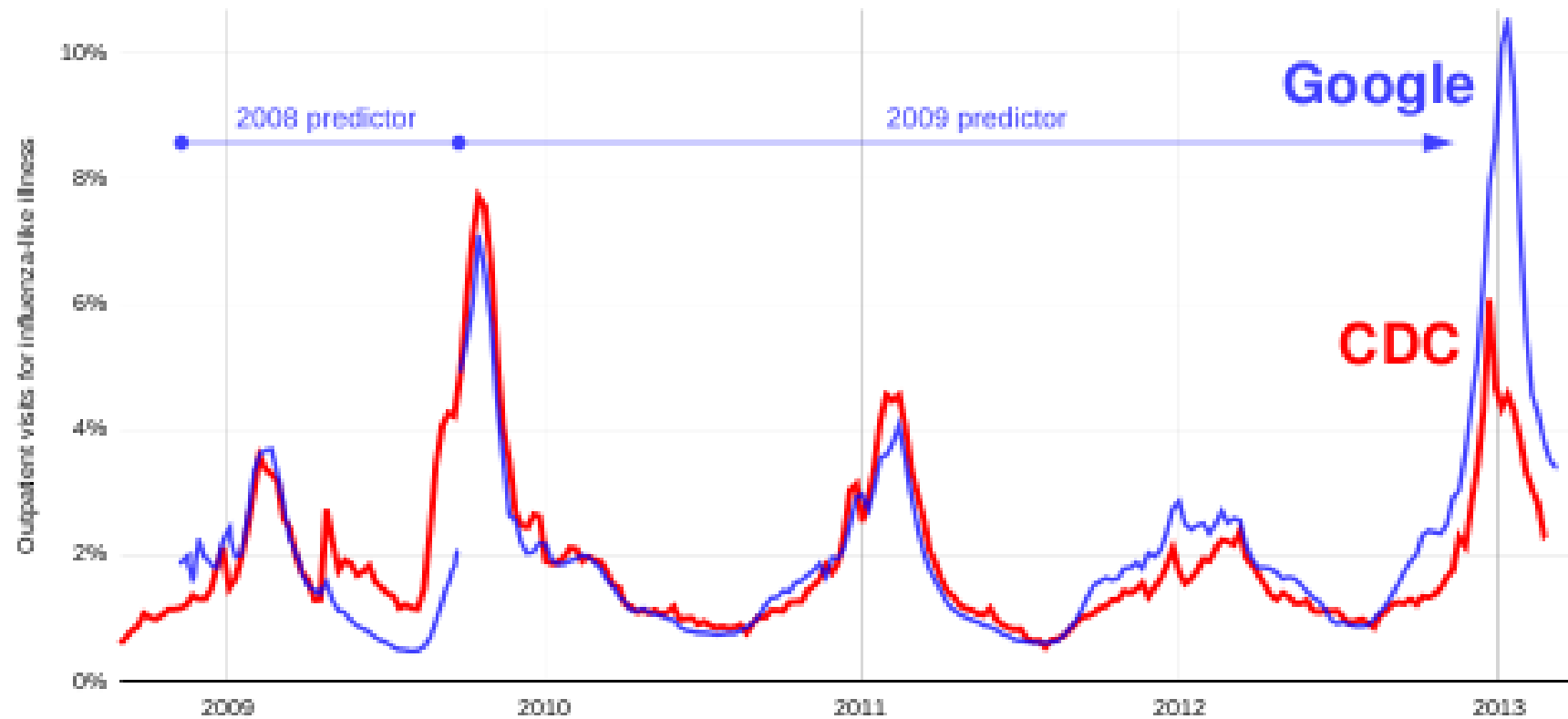




# 구글 GET의 신화



## Second divergence in 2012–2013 for U.S.



# 왜 구글의 독감 추적은 실패했을까?

**“구글 플루 트렌드를 모든 의사결정과정에서 적용했다!”**

- 1. 플루와 관계 없지만 특정 계절에 많이 언급되는 단어에 취약**  
**Ex. 고교 농구**
- 2. 검색어 정보에 기반한 효과적인 의료 정보 제공 부가서비스 시작**  
**-> 사람들의 특정 단어 검색 늘어나 GET 추적 경로 이탈**
- 3. 그 해 유행했던 플루에 대한 공포**  
**-> 실제 증상을 가진 사람에 비해 검색 횟수 높아져…….**

# 인과관계와 상관관계 사이의 긴장

010001000

**BIG  
DATA**

101011010

**≠**



과거에도 그랬으니 미래에도 이럴 것이다

과거 설명과 미래 예측은 근본적으로 다르다

**Overfit 가능성 농후!!!**

# GET의 연구는 과연 정확한가?



**“실리콘밸리식의  
아마추어 사회 과학이 만들어 낸  
예상했던 실패”**

**2011년에서 2013년 사이의 108주 중 100주  
GET의 예측치가 실제 값을 넘어섰고  
심할 때는 두 배나 되는 경우도 있어…….**

# 빅데이터는 건초 더미 속에서 바늘 찾기!!



GET의 목표

측정, 개념의 타당성, 일관성, 자료 간의 비독립성  
힘들어요

-> 빅데이터와 기존의 자료 합쳐서



# 필요 기술

REVERSE ENGINEERING

넷플릭스의 예시를 중심으로

- **ICT 기술(빅데이터, 인공지능, 슈퍼 컴퓨팅)**

- **분석 기술**

**데이터 마이닝, 기계 학습, 자연어처리, 패턴 인식 등**

- **표현 기술**

**R, Python 등의 프로그래밍 언어**



# REVERSE ENGINEERING

역공학

# Reverse engineering?



- 제작과정을 역행하는 과정
- 기존 소스코드 -> 오류, 미완성, 비즈니스 룰, 구조 등

# 예시 1 - MICrONS 프로젝트



1. 뇌 전체를 지도로 만든다
  2. 기계학습과 인공지능에 응용한다
- (가설)Bayesian Engine

미국 정부 주도,  
뇌의 알고리즘 규명 후 컴퓨터에 응용한 인공지능

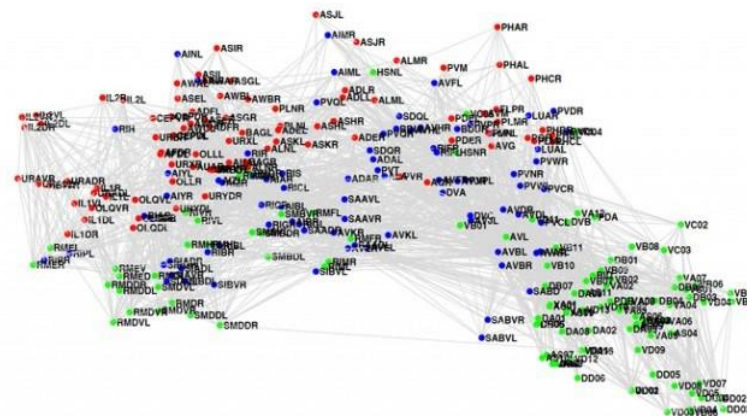
# 예시 2 - 예쁜꼬마선충 시뮬레이션



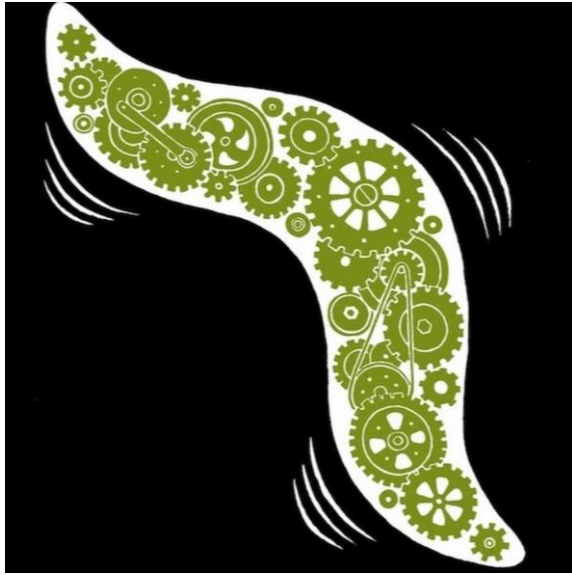
전체 게놈 서열이 밝혀진 첫 생명체

특정 신경에 어떤 자극을 줄 때 무슨 행동 하는지 다 안다!

-> 예쁜꼬마선충 시뮬레이션 제작  
[Open worm]



# 사회물리학 또한 마찬가지로 과정



왜 그렇게 되는지는 몰라~

하지만 현실세계에서도 이렇게 작동해~

# 넷플릭스의 예

컴퓨터 활용 이전에도 이후에도  
사회물리학 이용한 대표적인 성공 기업

# 넷플릭스 ver1.0 = DVD 배달 업체



우편을 통한  
자동배송, 자동 반납 서비스

# 블록버스터를 이긴 넷플릭스

*"There can be only one."*



- 땅 값 낮은 곳에 물류센터 (컨텐츠 투자 비용 확보)
- 판권 투자 (새 영화 & 오래된 영화)
- 물류 요충지에 창고, 우체국과 벌크요금 합의
- 손익 분기점 20편
  - 유통 구조로 손해 예방





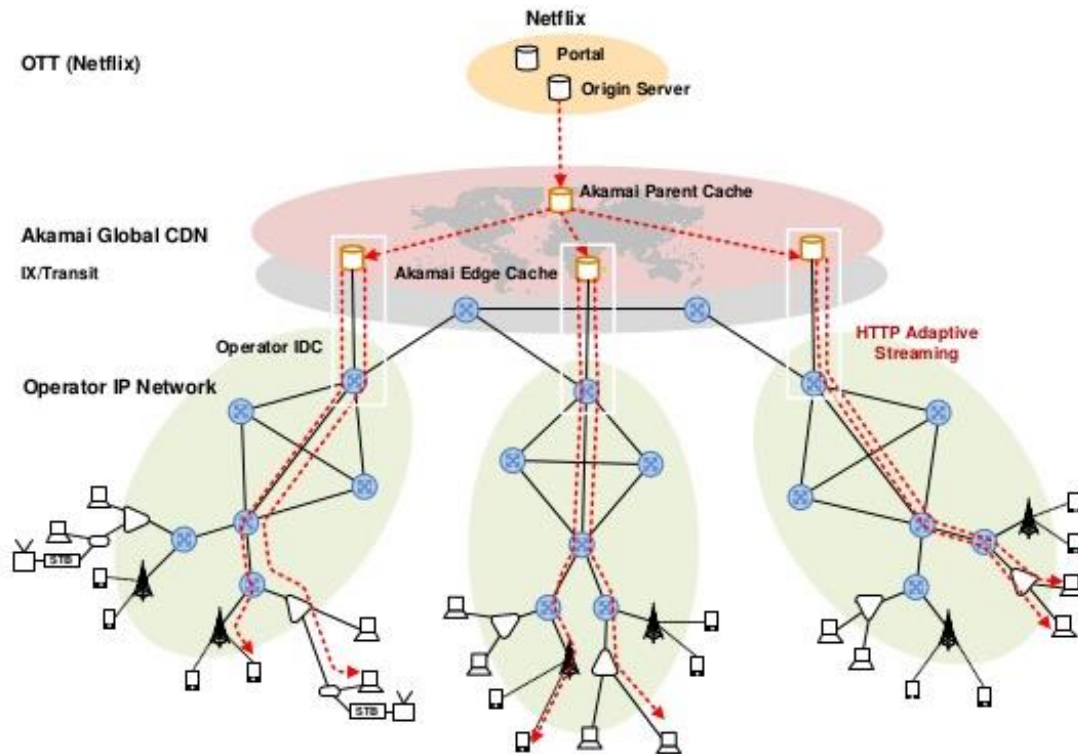
**Q. 블랙버스터 사에 맞선 비결은?**

**A. 부족한 콘텐츠와 자금을  
분석으로 충당했습니다~**

# Netflix streaming service (2007)

## Case Study: Netflix over Akamai CDNs

NETMANIAS  
www.netmanias.com



- 스트리밍 서비스 시작!!
- CDN(Content Delivery Network)  
컨텐츠를 사용자 근처지역에 캐시
- 스트리밍용 하드웨어 직접 설계
- 일부 ISP또한 무료로 넷플릭스에 저장 시스템 제공

# NETFLIX



어떤 영화를 좋아할까?



어떤 영화에 투자할까?

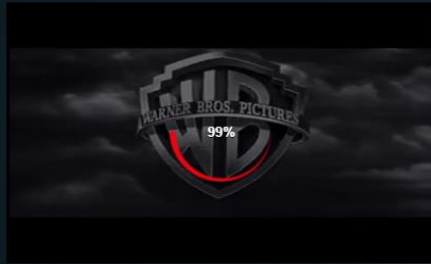
# 사람들은 어떤 영화를 좋아할까

별점을 넘어선 개인화 시스템

수많은 태그

A/B 테스트

영화 본 뒤 연이은 추천



인셉션



**추천 콘텐츠**  
오른쪽의 콘텐츠를 추천해 드립니다.

뒤로 가기



# Netflix Prize - 1백만달러 드려요!!



가장 효율적인 별점 추정 시스템 코딩 경진대회  
하지만 시청을 10%밖에 올리지 못함.

# 협업 필터링

사용자 기반 알고리즘



시연 : 겨울왕국 5점 컨저링 2점 라퐁젤 4점  
채은 : 겨울왕국 2점 컨저링 4점 라퐁젤 3점  
지영 : 겨울왕국 4점 컨저링 1점 라퐁젤 ???



아이템 기반 알고리즘



# 별집을 넘어서

- 개인화 시스템
- 보기패턴이 비슷한 사람들 묶기
  - 요일, 시간대, 장치, 위치
  - 페이스북 친구와 연동
- A/B 테스트 기법
- 엔딩 크레딧이 뜨면 추천 시스템 작동
- 다음 에피소드 자동 시작

I love 다큐  
I love 외국영화

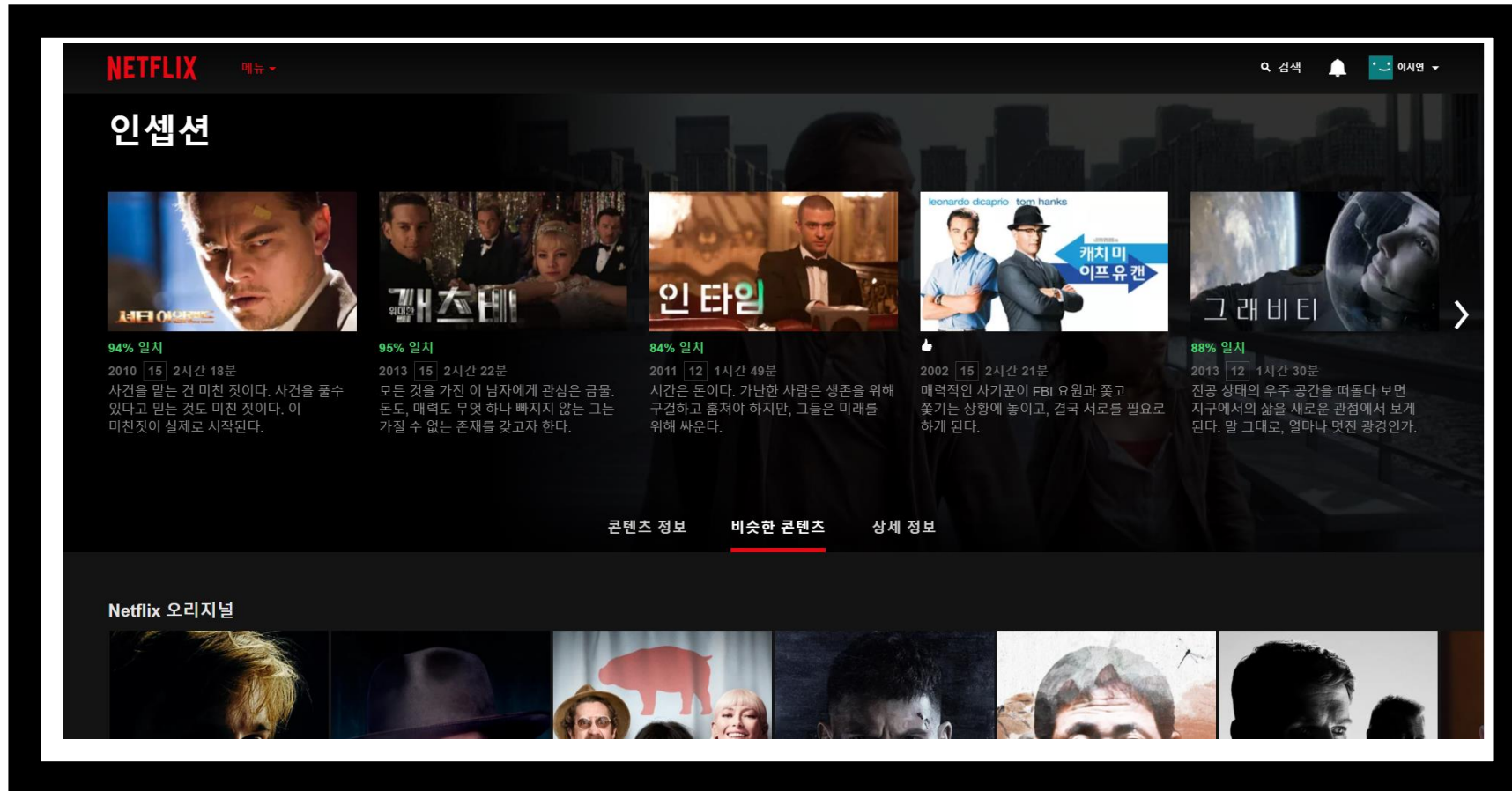




# 넷플릭스의 정보 수집에 동의하십니까?

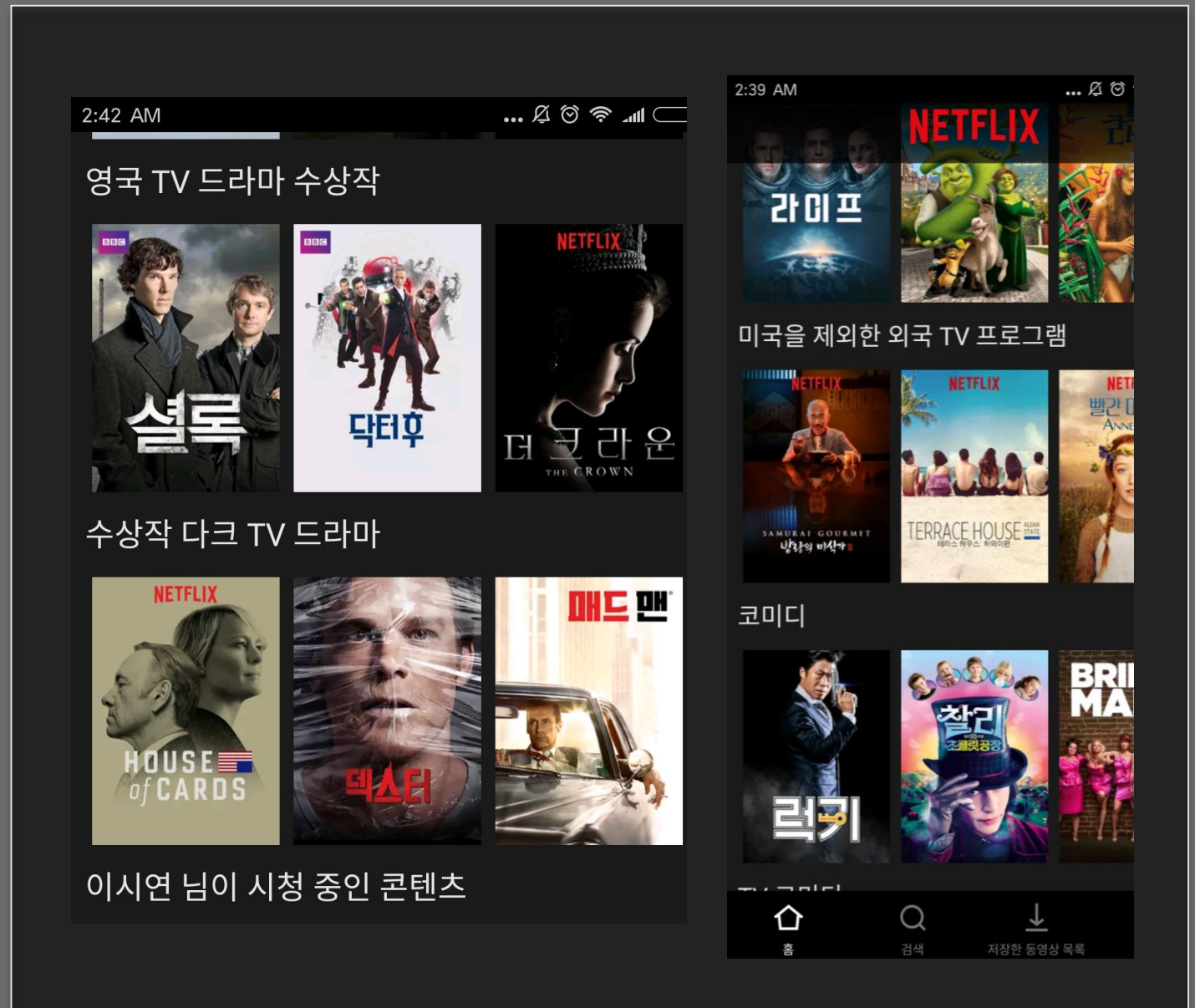
- ✓ 특정 영화 시청 중 일단정지, 되돌리기, 빨리가기 하는 지점
- ✓ 시청한 요일, 날짜, 시간
- ✓ 시청한 장소(주소)
- ✓ 시청에 사용한 기기
- ✓ 시청을 중단한 지점
- ✓ 시청 후 사용자가 준 별점
- ✓ 사용자이 영호 검색 내용
- ✓ 영화를 고르는 동안 하는 행위

# 개인화를 통한 영화 추천



# 영화를 구분하는 76,897 방법

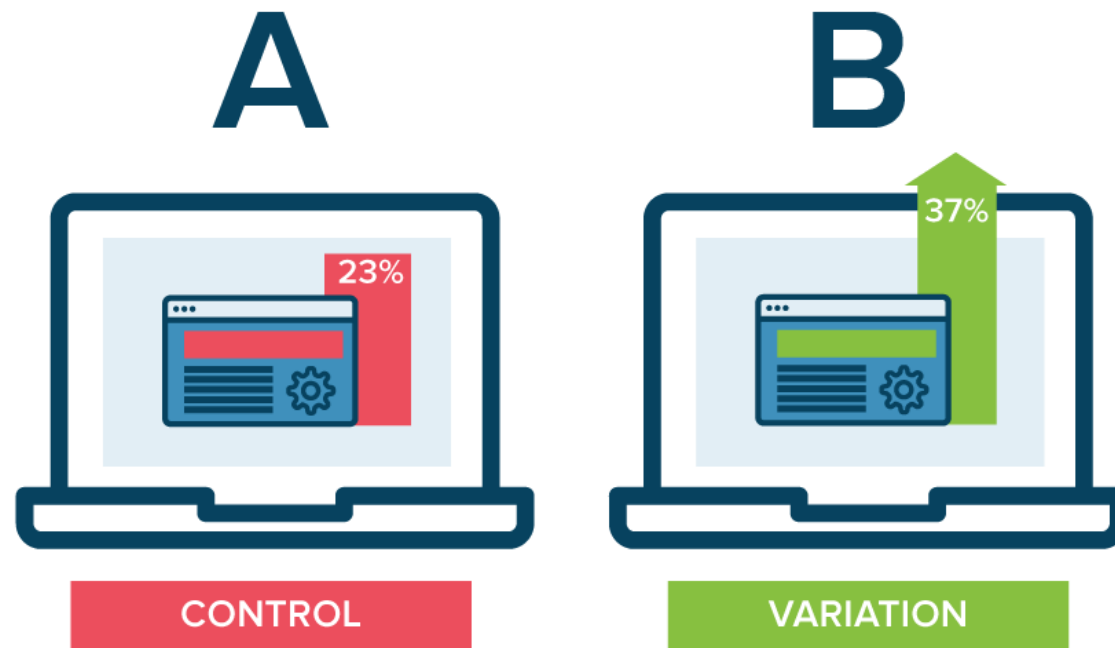
- 넷플릭스의 영화장르는 리버스엔지니어링 통해 밝혀진 것만 76897개.
- 단순한 분류 아닌 개인화 추천



# 넷플릭스식 이름짓기

- 악마 같은 아이가 나오는 컬트 공포 영화  
(Cult Evil Kid Horror Movies)
- 유럽 배경의 60년대 영국 공상과학/판타지물  
(British set in Europe Sci-Fi & Fantasy from the 1960s)
- 비평가들에게 호평받은 감동적 패배자 영화  
(Critically-acclaimed Emotional Underdog Movies)
- 지역 + 수식어 + 장르 + 원작 + 배경 + 시대 + 주제 + 나이
- 시각적으로 아름다운 시립대 배경의 물리학과에 대한 감동적 액션 영화

# A/B 테스트



인과관계 속 상관관계 제거

**투자도 잘하는 넷플릭스**

# 하우스 오브 카드

1. 원작 소설 영국에서 인기가 좋았다.
  2. 원작을 보 사람은  
데이비드 핀처 감독과 케빈 스페이시가 나오는 영화도 본다.
  3. 데이비드 핀처 영화 '소셜 네트워크'  
많은 시청자가 처음부터 끝까지 봤다.
- 이 영화 대박이다! 판권에 과감한 투자

# 아마존은 망했고 넷플릭스는 살았다

- 아마존의 알파 하우스는 망함
- 시즌1 13편을 일시에 보여줬다  
(주말내내 몰아본 시청자 67만명, 직장 결근까지 생길 정도)
- 200만명의 신규 가입자(미국)  
그 외 백만명



**현황 및 관련 사업**



**2010년 인구 총조사**



**인구성장률  
노동인구  
유소년 및 노령 인구 변화 예측  
시군구 별 인구 이동변화 예측**



**국민기초생활보장  
기초노령연금  
영유아 복지  
노인 일자리 등  
사업별 수급자 및 예산 변화 전망**

# 해외 기업



RateMyDrive



고객 임신여부  
- 프로모션

## Ship Before They Buy



Amazon.com plans to ship you things before you even buy them. Using predictive analytics, the online retailer will guesstimate your next purchase.

# 국내 기업



영화 추천 스트리밍 서비스

## Macro Trend와 Micro Trend를 조합한 남/녀 각 9개 Trend Code

M			W		
<b>Rookie</b>  새롭고 다양한 분야에 관심이 많은 프레스맨	<b>LOEL</b>  패션 및 명품 브랜드에 관심이 많으며 외모가꾸길 즐기는 센스남	<b>Friend Daddy</b>  자녀와의 친밀감을 중시하는 친구처럼 가까운 아빠	<b>it-Girl</b>  최신 유행 및 트렌드를 선도하는 패셔니스타	<b>Prima Donna</b>  문화와 여가를 즐기는 전문직 싱글 여성	<b>Trend Setter</b>  럭셔리한 삶을 추구하는 세련된 감각의 여성
<b>Smart Saver</b>  합리적 가격을 중요시하는 계획적인 소비를 즐기는 플랜맨	<b>Mr. Routine</b>  하루하루 최선을 다하며 소소한 행복을 추구하는 가장 standard 한 직장인	<b>BOBOS</b>  일과 여가를 즐기며, 독특한 소비감각을 지닌 여유로운 남성	<b>Rudy</b>  자기계발에 적극적인 젊은 감각의 여성	<b>Alpha Mom</b>  자녀교육에 매진하는 똑똑한 엄마	<b>Queen House</b>  경제관이 뚜렷하여 가족을 적극 보살피는 내조의 여왕
<b>Bravo Life</b>  사회적 기여에 관심이 많고, 젊고 댄디한 감성을 잃지 않은 남성	<b>Realist</b>  건강과 웰빙, 일과 여가의 균형을 추구하는 이성적 남성	<b>Gray Gentleman</b>  필수적 소비에 집중하며 삶의 질을 중시하는 시니어 남성	<b>줌마엘라</b>  외모와 건강에 관심이 많은 사교적인 여성	<b>Grace Woman</b>  레저와 여가를 즐기며, 기부활동에 적극적인 여성	<b>Silver Lady</b>  건강을 유지하며 삶 자체를 즐기는 시니어 여성

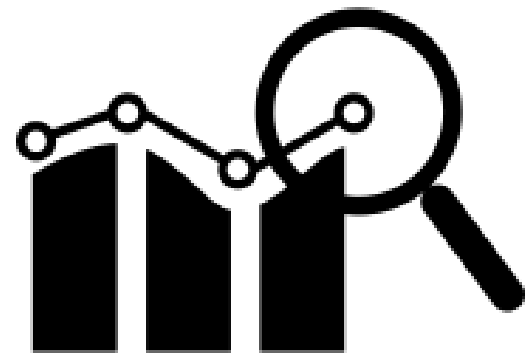
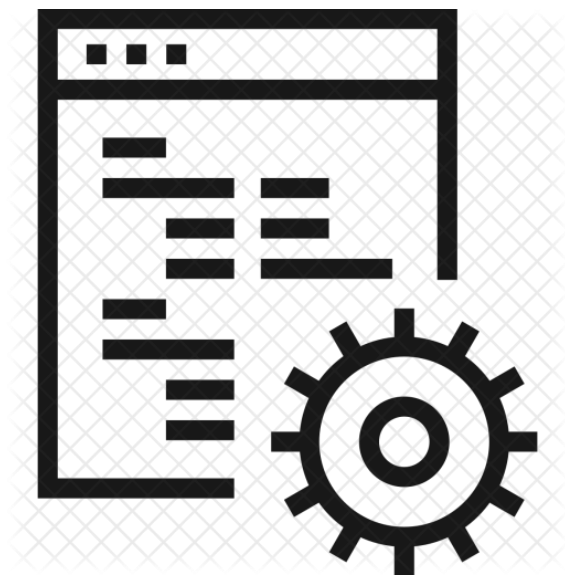
# 국내 활용도는 아직 낮은 편

- 중견기업 빅데이터 도입률 9.6%
- 종업원 100명 이상 도입률 4.3%
- 데이터 절대적으로 부족
- 개인정보 보호, 데이터 개방 거부감

# 사회물리학 전문 기업은 없다

- 하지만 넷플릭스, 구글 포함 다수 기업들  
사회물리학을 통한 기업의 발전 보여줌
- 기술과 경영, 마케팅 등 모든 영역을 총괄해서  
꼭 필요한 학문





**그러니까 우리는**

# 참고 문헌 및 출처



# 참고 문헌 및 출처

- 물리학은 사회현상을 설명·예측할 수 있는가?, 사이언스온, 조향현, 2010.12.09, [http://scienceon.hani.co.kr/?document\\_srl=33737](http://scienceon.hani.co.kr/?document_srl=33737), 2017.09.15
- 게임이론:사회과학 방법론의 통일, 물리학과 첨단기술, 김영세, 2015년 1/2월 제24권 12호
- 우리 일상의 게임이론, 위의 책, 김범준, 정형채
- 사회물리학, 위의 책, 정우성, 2011년 5월 제20권 5호
- 빅데이터 분석의 국내외 활용 현황과 시사점, 최재경, KISSTEP InI 제 14호 (2016.06)
- 모사현실을 통한 미래사회 탐색과 예측, 소아영, 융합연구정책센터 2017 SEPTEMBER vol.86 (2017.09.04)
- 빅데이터와 사회과학하기: 자료기반의 변화와 분석전략의 재구상, 한신갑, 한국사회학 제49집 제2호(2015년), pp.161~192
- [주간기술동향] 뇌의 리버스 엔지니어링, 미국의 워신을 건 AI 개발 국가 프로젝트 (2016.06.28) 정보통신기술진흥센터
- 넷플릭스는 어떻게 작동하는가, 네이버레터, 박상현, <http://nter.naver.com/naverletter/110357>, 2017.12.01