

## 『빅데이터 인문학: 진격의 서막』 4-7장

발제 : 신정아

### 4장. 사람은 어떻게 유명해지는가?

엔그램을 이용해 명성을 탐구하는 것은 아주 흥미롭고 당황스럽고 재미 다. 그러나 엔그램 주변에는 어둠도 도사리고 다. 어떤 비밀도 이것보다 더 어둡지는 않을 것이다. 바로 극도로 사악한 행동보다 더 효율적으로 명성을 만들어내는 것은 없다는 점이다. 우리는 명성으로 가는 가장 확실한 길이 사람을 죽이는 일인 세계에 살고 다. 이것이 뜻하는 바가 무엇인지는 우리 모두가 깊이 생각해보아야 한다.[104]

#### 명성은 어떻게 찾아오는가

- 명성은 황홀하지만 어떻게 작동하는지 숨겨져 있는 것[114]
- 명성의 여러 측면을 측정하는 것은 어렵지만, 크기를 재는 것은 가망 있음
- 사람들이 얼마나 자주 당신의 이름을 언급하는가에 달려 있음

#### 거의 유명하면 충분히 유명하다

- 엔그램으로 측정하는 것은 명성fame 그 자체가 아니라 단순화된 명성의 복제품으로 여기서는 ‘미영성 phame’이라고 칭함[118]
- 명성이란 에너지나 생명처럼 직관적으로 파악하는 일상생활의 개념이지만 정의하기가 극도로 어려움
- 영국의 유명작가 찰스 디킨스의 ‘미영성’
  - : 『올리버 트위스트』(1837), 『크리스마스 캐럴』(1843), 『데이비드 카퍼필드』(1849), 『두 도시 이야기』(1859), 『위대한 유산』(1860) 등의 베스트셀러를 꾸준히 내놓으면서 유명해짐
  - : 작품이 유명해질 때마다 작품 속 문구가 인기를 끌게 됨
  - : 1870년 디킨스의 사망은 그의 천재성에 관한 새로운 논평을 낳았고, ‘미영성’은 하늘로 치솟음. 사망 후 수십 년 간 디킨스를 언급하는 빈도는 절정에 달함
  - : 1900년부터 서서히 줄어들면서 디킨스의 ‘미영성’은 완만하게 시들해지고 있음. 약 1세기에 걸쳐 ‘미영성’이 이어짐
- 디킨스의 측정된 ‘미영성’의 오류 가능성은 없을까? [120-121]
- 통계학자들은 측정도구가 만들 수 있는 오류를 두 가지 유형으로 구분함
  - 1) 무작위 오류 : 측정대상이 변화하지 않는데 변동이 생기는 경우. 이러한 오류들은 ‘미영성’의 작은 붕 우리와 계곡의 형태로 나타나지만 일반적으로 참값true value에 가까이 머무름
  - 2) 시스템 오류 : 부풀리거나 줄이는 방향으로 측정을 왜곡함. 예를 들어 어떤 사람의 이름이 쓰인 사례를 찾는데, 모든 참고자료 가운데 일부분만을 포함함. 디킨스의 빈도를 추적함. 찰스 디킨스의 경우 ‘디킨스’ 또는 ‘찰리’ 또는 ‘시-머니 C-Money’라고 부르는 사례를 놓치게 됨. 또한 누군가가 디킨스를 가리키며 그가 쓴 유명한 구절을 인용하거나, 마술사 데이비드 카퍼필드의 마술에 감탄하거나 심지어 ‘메리 크리스마스’라는 구절을 쓸 때도 우리는 그것을 놓치게 됨
- 시스템 오류에는 부정 오류false negative와 긍정 오류false positive가 있음
- 부정 오류는 찾아내고자 했던 것을 소홀히 했을 때 발생하는 유형

→ 포함시켜서는 안 될 것을 포함시킬 때 나타나는 유형

- 3) 표본 오류 : 1936년에 태어난 세 사람의 예에서 알 수 있음. 헐리우드 배우인 로버트 레드포드, 체코 슬로바키아 초대 대통령 바츨라프 하벨, 뉴욕대 교수 캐럴 길리건 중 가장 ‘미영성’이 높은 사람은 캐럴 길리건. 페미니스트로서 획기적인 업적을 쌓았고, <타임>이 뽑은 가장 영향력 있는 미국인 명단에 오른 인물. 캐럴 길리건을 언급한 책들도 많아서 바츨라프 하벨이나 로버트 레드포드보다 더 많이 언급됨
- : ‘미영성’과 명성이 같은 것이라면 가장 유명한 사람은 캐럴 길리건. 그러나 이 결과는 책에서 언급된 결과이기 때문에 전체를 포괄하지 못함. 텔레비전, 신문, 인터넷 등의 평가를 포함한다면 하벨과 레드포드가 길리건을 큰 차이로 앞질렀을 것 → 표본 오류
- : 길리건에게 불공평하게 이득을 줌. 길리건은 명성보다 ‘미영성’이 더 높음
- : 표본 오류를 줄일 수 있는 정교한 데이터 세트를 포함시키는 것이 필요함 (타블로이드 신문, 잡지, 학술논문 등)

#### 명성을 질병처럼 다루기

- 1930년, 크리스티안산Kristiansand에서 활동했던 의사 안드보르드가 행한 코호트cohort 분석[125]
- 1887년부터 1891년 사이에 태어난 아이의 1퍼센트 이상이 태어난 지 1년 이내에 결핵으로 사망. 그중 11~15세가 절반 가까이 차지함. 수십 년간 유행병이 창궐하는데 평균 연령은 올라가는 현상이 나타남
- 전체 인구의 발병률 대신 같은 시기에 태어난 사람들의 집단인 코호트로 전체 인구를 나누어서 결핵이 한 세대에게 미치는 영향을 분석함
- 그 결과, 태어난 연도에 상관없이 5-14세 사이, 혹은 20-24세 사이에 결핵에 감염될 가능성이 가장 높은 것으로 나타남. 결핵이 젊은이들에게 주요한 질병이었다는 것을 증명
- 결핵에 관한 연구처럼 명성에 관한 연구도 온갖 종류의 특정 세대에 국한된 효과들 때문에 혼란에 빠짐. 명성도 코호트 방식을 사용해야 함. 명성도 질병처럼 다뤄야 함

#### 명예의 전당

- 1800년에서 1950년 사이에 태어난 사람들을 대상으로 매년 해당 연도에 태어난 사람 가운데 가장 유명한 150명의 명단을 작성함(명단 목록 : 131~134쪽)
- 수십만 명을 대상으로 성과 이름 전체(예를 들어 마크 트웨인)가 언급된 빈도를 계산함
- 해마다 한 명씩 그러나 한 눈에 봐도 이 명단은 우리 아이들에게 넘겨줄 역사 서술의 기반이 될 수 없음. 그 이유는 해마다 가장 유명한 한 명씩 뽑은 150명 가운데 열여섯 명만이 여성이고 절대 다수가 백인 남성. 목록 자체가 깊은 편견을 안고 있음[135]
- 목록의 편견은 역사적 기록이 갖는 본질적 편견임. 어느 누구도 표본 추출의 편견에서 자유롭지 못함
- 역사는 편파적일 수 있어도 통계는 그럴 수 없음
- 엔그램 데이터는 우리가 무엇을 잘못하고 있는지 좀 더 명확한 그림을 보여줌으로써 우리가 얼마나 깊은 편견에 빠져 있는지를 가능하게끔 함
- 우리가 과거의 편견을 더 잘 기억할 수 있게 되면 그것을 되풀이하는 데에서 벗어날 수도 있을 것

#### 더 빨리, 더 많이 유명해지는 만큼 더 빨리 잊힌다

- 1871년에 태어난 가장 유명한 50명을 들여다보니 하나의 평균적 형태가 드러남[138]
- 데뷔, 기하급수적 성장, 절정, 점진적 쇠락을 보여주는 곡선 형태

- 데뷔한 연도, 기하급수적 성장의 속도, 정점을 지난 뒤 쇠퇴하는 속도 등 세 가지 매개변수parameter로 측정해보면 모든 집단은 대략 태어난 지 4분의 3세기쯤 정점에 도달함
- : 젊은 나이에 유명해지고, 명성이 높아지는 속도도 빨라짐. 1800년생 집단이 명성이 두 배가 되는 데 약 8년이 걸렸고, 43세에 데뷔하여 75세에 절정에 이르면 명성이 네 배 더 커짐
- : 1950년생 집단의 경우, 명성이 두 배로 높아지는 기간이 훨씬 짧아져 3년
- : 곡선 형태는 각각의 코호트별로 같아 보이지만 젊은 집단이 높은 집단에 비해 훨씬 더 유명함. 젊은 코호트는 괴롭힘을 당할 가능성이 더 높음
- : 정점을 지난 뒤 쇠퇴하는 속도는 1800년의 반감기는 120년, 1900에는 71년으로 떨어짐
- : 유명해지는 만큼 더 빨리 잊혀짐[141]

## 지난 200년 동안 가장 유명한 사람

- 우리의 접근법은 명성과 오명을 구분하지 못함
- 엔그램 데이터는 이름 앞뒤에 어떤 단어들이 오는지 충분한 맥락을 제공하지 않기 때문에 긍정적인지 부정적인지 판단할 수가 없음
- 지난 2세기 동안 태어난 사람들 가운데 가장 유명한 10명
- : 아돌프 히틀러 / 카를 마르크스 / 지그문트 프로이트 / 로널드 레이건 / 이오시프 스탈린  
블라디미르 레닌 / 드와이트 아이젠하워 / 찰스 디킨스 / 베니토 무솔리니 / 리하르트 바그너
- 히틀러가 맨 위. 대량 학살자가 세 명이나 포함됨
- 살인과 명성은 서로 연결되어 있음. 이러한 끔찍한 현상이 가져오는 역설은 그 사건이 있기 전에는 전혀 알려지지 않았던 살인자들이 미디어 태풍의 중심이 됨
- 1940년까지는 히틀러나 스탈린이 등장하지 않지만, 1950년부터 등장함
- 한편으로 이런 종류의 뉴스는 무슨 일이 일어났는지 사람들이 알아야 하기에 중요함. 그러나 이렇게 생겨난 관심이 살인자들에게 동기로 작용할 수도 있음 (존 레논 살인자 호프먼처럼) [148]

- 엔그램을 이용해 명성을 탐구하는 것은 아주 흥미롭고 당황스러움
- 엔그램 주변에는 어둠이 도사리고 있음. 어떤 비밀도 이것보다 더 어둡지는 않을 것
- 극도로 사악한 행동보다 더 효율적으로 명성을 만들어내는 것은 없음
- 우리는 명성으로 가는 가장 확실한 길이 사람을 죽이는 일인 세계에 살고 있음
- 엔그램 데이터의 힌트 또는 가능성
- : 히틀러보다 먼저 태어난 덕분에 1880년부터 1940년까지 가장 유명한 사람은 ‘찰스 디킨스’
- : 다정하고 사랑스러운 희극작가, 찰스 디킨스. 평화와 전쟁. 최고의 시절이자 최악의 시절.

## 5장. 침묵의 소리: 빅데이터가 말하는 억압과 검열의 역사

독일어로 판 책들을 이용해 계산 이 엔그램들은 샤갈 및 그의 동시대인에게 나치가 가한 억압이 만들어 낸 효과를 똑똑히 보여준다. 1936년에서 1943년 사이에 마르크 샤갈이라는 이름은 우리가 가진 독일어 책 기록에서 딱 한 번 등장한다. 나치는 샤갈을 죽이지 못했다. 그러나 그를 지워버리는 방법을 찾았다.[152]

- 책에 반영된 수백 만 명의 목소리는 문화와 역사에 관한 매혹적인 이야기를 들려줌

- 그러나 모든 사람의 목소리가 기록되는 것은 아님
- 때로는 사라진 목소리들의 침묵이 모든 것을 들리게 않게 할 수도 있음
- 1880년에 태어난 헬렌 켈러, 가장 영향력 있는 작가이자 정치적 행동가, 인간승리의 상징으로 추대됨. 그러나 나치는 1933년 독일을 접수한 후 독일 사람들, 독일 문화를 통제하기 시작함. ‘비(非) 독일적 정신 un-German spirit’을 반영한다고 생각하는 책들을 탄압 → 도서관과 서점에서 책들을 수거, 독일 전역에서 분서(焚書)의 불길에 휩싸임
- ‘사상을 죽일 수 있다고 생각한다면 역사가 여러분에게 아무 것도 가르치지 못한 것이다’라는 켈러의 주장으로 나치의 분서는 국제적 공분을 삼
- 켈러가 여론의 재판정에서 승리했지만, 과연 그녀가 옳았을까? 사상을 죽이는 일이 정말로 불가능할까?

## 샤갈의 명성

- 모더니즘 운동의 개척자. 20세기 중반의 가장 선두적인 화가
- 1917년 러시아 혁명 직후, 서른의 나이에 러시아의 시각예술 전반을 감독하는 정치위원을 제안 받음
- 그러나 계속된 전쟁과 기근으로 러시아인들의 삶이 파헤쳐지자 파리 이민을 택함
- 1923년 파리에 도착한 후 자신의 명성과 평판에 충격을 접하면서, 고향에서 자신의 이미지가 ‘희미해질까 봐’ 두려워함[156]
- 샤갈의 명성이 책 속의 언급으로 이어진 범위 안에서 명성과 평판을 분석함
- 이민의 효과로 1923년 이후 프랑스어 책에서는 언급이 많아지고, 러시아어 책에서는 줄어듦[156]

## 퇴폐 미술전

- 1920년대 독일은 미술의 천국. 다다, 바우하우스, 표현주의, 큐비즘 등이 독일에 뿌리를 둠
- 아돌프 히틀러는 새로운 동향을 강력하게 반대하여 제국문화부 장관 요제프 괴벨스를 통하여 나치가 퇴폐적이라고 여기는 예술품들을 압수함
- 유대인 초현실주의 샤갈을 정조준함. 이와 동시에 조르주 브라크, 폴 고갱, 바실리 칸딘스키, 앙리 마티스, 피터르 몬드리안, 파블로 피카소 등 오늘날 세계적으로 유명한 현대 미술가들의 작품 상당수를 비롯한 수천 점의 ‘퇴폐’ 작품들이 압수됨
- 나치는 아방가르드 작품을 압수하고, 작품 활동을 방해하는 것은 물론이고, 평판도 떨어드리고 싶어함
- 1937년 <위대한 독일 미술전>과 <퇴폐 미술전>을 동시에 개최
- <퇴폐 미술전>을 통해 나치는 현대 미술이 도덕적으로 파산했고, 지독히 상업적이며, 납세자들의 돈을 낭비한다고 묘사함. 대중의 분노와 경멸을 조장함.
- <퇴폐 미술전>은 넉 달 동안 200만 명, 하루 평균 1만 7,000명의 관람객을 동원. 미술사에서 전무후무한 기록을 달성. 전대미문의 인기 전시회가 됨. 이후 독일에서 현대 미술가가 되는 것은 사실상 불가능해짐
- 샤갈은 독일 문화에서는 빠르게 지워졌지만, 프랑스에서 한동안 안전하게 거주하다, 1940년 프랑스 함락 이후 생명의 위협을 느낀 후 가족과 함께 미국행을 택함
- 독일어로 출판된 책을 분석한 엔그램 데이터에서 1936년에서 1943년 사이에 샤갈의 이름은 단 한 번 등장함. 나치는 샤갈을 죽이지는 못했지만 그를 지워버리는 방법을 찾아냄[164]

## 나치의 분서 정책

- 나치 정권 하에서 독일의 주요 학생 연합체인 독일학생연합은 독일문화에서 바람직하지 않은 신념들을 청소하는 전국가적 운동을 개시함

- 독일학생연합에서는 도서관에 ‘비독일적 정신’을 반영한 작품이 있다고 믿음
- 1931년 나치당에 합류한 사서 볼프강 헤르만은 베를린의 도서관들을 검사하기 위해 설립된 ‘정화위원회’에 배치됨. 헤르만은 자신이 ‘독일의 매음굴’이라고 이름붙인 책들의 목록과 정치인, 문필가, 철학자, 역사가를 비롯한 모든 부류의 작가에 대한 개별 목록을 독일학생연합에 제공함
- 1933년 5월 10일, 무장한 독일학생연합은 헤르만의 목록을 바탕으로 서점, 공공 도서관, 학교를 습격해서 수만 권의 책을 불태움. 이후 5월 말까지 독일 전역에서 분서가 일어남[166]
- 5월의 분서를 시작으로 나치는 기나긴 공격을 시작함. 헤르만의 목록은 1933년 작가 500명을 시작으로, 1938년 수천 명으로 늘어남
- 엔그램을 사용하여 검열 운동의 효과에 대한 통찰을 얻을 수 있음. 헤르만의 블랙리스트에 이름이 오른 지식인들과 나치 정권과 연계된 인사들이 보인 양상은 극명한 대조를 이룸
- 헤르만의 전투 효과는 분야별로 다르게 나타남. 철학·종교 서적 작가들의 명성은 1/4 수준으로, 정치 관련서를 쓴 작가들은 절반으로 떨어짐. 반면 역사가들에 대한 효과는 10퍼센트 정도만 낮아짐[167]

## 사라진 이름들

- 소비에트사회주의연방공화국 건국의 도화선이 된 러시아 혁명. 레닌의 뇌졸중 이후 권력투쟁이 시작됨
- 레닌과 볼셰비키를 이끈 트로츠키가 레닌의 후계자로 예상되는 가운데, 혁명의 세 영웅 오시프 스탈린, 그리고리 지노비예프, 레프 카메네프가 ‘트로이카’ 동맹을 맺고 트로츠키를 공격함
- 1925년 소련의 지도자가 된 스탈린은 절대 권력을 위해 지노비예프와 카메네프를 숙청함
- 트로츠키 역시 사형 선고를 받은 후 멕시코에서 살해됨
- 스탈린은 정적을 죽이는 것에서 그치지 않고 자신이 유일하고 대표적인 혁명 영웅으로 남기 위해 처형 이후 반 세기 동안 트로츠키, 지노비예프, 카메네프, 다른 셀 수 없이 많은 사람들의 공헌을 최소화함
- 엔그램 데이터를 보면, 세 사람의 명성은 대숙청 이후 가파르게 떨어짐
- 스탈린 사후 흐루쇼프가 1956년 공개적으로 대숙청을 비판했지만 세 사람의 평판은 부분적으로만 갱생함. 여러 세대가 흐른 뒤 1980년대 후반 미하일 고르바초프 시대에 가서야 평판이 올라감[170]

- 2차 세계대전 후 미국에서는 공산주의에 대한 불안이 일어남
- 공산주의자들을 조사하기 위해 1945년 ‘하원 비미(非美) 활동위원회’라는 특별 상설위원회를 설치
- 영화 산업이 대외 선전의 비밀스런 원천이 될까 우려하여 공산주의자와 연계됐다고 의심되는 배우호적 배우와 감독들을 상대로 그들이 알고 있는 것을 밝히라고 압박
- 당시 증언을 거부한 엘바 베시, 허버트 비버먼 등 유명 영화인 열 사람을 이후 ‘할리우드 텐Hollywood Ten’이라 칭함
- 할리우드 주요 제작자 48명은 반공산주의자 평판에 기대어 스스로 공산주의자가 아니라고 법정에서 선언하지 않으면 자신들의 스튜디오에서 작업할 수 없다고 선언함. 이들은 나중에 할리우드 텐뿐만 아니라 다른 많은 이들까지 미국에서 일자리를 얻을 수 없도록 블랙리스트를 작성함
- 할리우드 텐 멤버들은 10년이 넘도록 메이저 스튜디오에서 제작된 영화에 이름을 올리지 못함
- 이후 1960년 블랙리스트가 깨지면서 추방된 자들이 다시 할리우드로 돌아옴[172]

- 1989년 천안문 광장의 학살 이후 중국 정부는 검열과 정보 억압을 통해 사건 이후 1년 이내 중국의 신문 가운데 10퍼센트 이상이 폐간되고, 출판사의 폐업도 이어짐
- 디지털 매체 역시 ‘중국의 거대한 방화벽’이라고 불리는 대규모 인터넷 검열의 일환으로 감시를 받음
- ‘천안문 광장’을 인터넷 검색해보면 세심하게 검열 처리된 결과만 확인할 수 있음

- 구글은 2006년부터 2010년까지 중국의 인터넷 봉쇄에 참여하는 데 동의함. 이후로는 협력을 중단함
- 결과적으로 중국의 많은 젊은이들은 1989년 6월 4일의 사건에 대해 거의 아는 바가 없음
- 반면 서구에서는 1989년 학살 이후 천안문에 대한 언급이 급증함[174]

## 검열을 자동으로 추적할 수 있을까

- 어디에서 일어나건 검열과 억압은 종종 특정 단어와 구절이 갑자기 사라지는 독특한 흔적을 남김
- 이런 어휘적 공백의 통계적 특징은 너무 강해서 우리가 숫자들(빅데이터)을 이용해 무엇을 억압받는지 파악하는 데 도움을 줄 수가 있음
- 예를 들어, 나치 독일 하에서 제3제국 시기인 1933~45년에 사각처럼 명성이 떨어진 사람을 찾는 것. 만약 누군가를 언급하는 빈도가 1920년대와 1950년대에 1000만 단어 당 1회였는데 나치 정권 기간에 1억 단어 당 1회로 내려갔다면, 이는 10분의 1의 감소를 뜻함. 이것은 그가 분명히 삭제되거나 어떤 방식으로든 억압받았음을 시사함. 반대로 100만 단어 당 1회로 올라갔다면 열 배나 상승한 것이므로, 정부 선전의 혜택을 받았을 것

## 억압과 검열을 넘어 백만 개의 통로로 스며들다

- 나치는 불과 몇 년이라는 짧은 기간에 위대한 사상가들을 지워버리기 위해 아주 먼 길을 감
- 현대 미술가들은 유럽에서 추방당하거나 강제로 은퇴했으며, 살해되기도 함
- 그러나 사상들은 살아남음
- 억압적인 정권이 사용한 전술의 의도치 않은 결과들에 주목할 필요가 있음. 샤를로테 살로몬의 사례
- 나치가 그것을 금지하고, 압수하고, 조롱하고, 작가들을 죽이는 잔인한 노력을 쏟아붓기까지 했는데도 현대 미술의 사상들은 죽지 않음
- 그것들은 진정으로 “백만 개의 통로로 스며들”었고, 살로몬이 <퇴폐 미술전>을 방문한 일처럼 예상치 못한 길에도 스며들었다. 살로몬은 죽었지만 그녀의 작품은 결국 “다른 사람들의 가슴을 뚫게” 했다.

## 6장. 기억과 망각의 속도

오랜 시간에 걸친 집단적 학습은 확연히 달랐다. 19세기 초기의 기술들은 영향력이 최고치의 4분의 1에 도달하기까지 65년이 걸렸다. 세기 전환기의 발명품들은 겨우 26년이 걸렸다. 집단학습 곡선은 10년마다 2.5년씩 줄어들며 갈수록 짧아지고 다. 사회는 점점 더 빨리 배우고 다.[184]

## 기억실험

- 19세기 독일의 철학자 헤르만 에빙하우스는 마음이 어떻게 작동하는지에 관심을 가짐
- 에빙하우스는 경험과 측정, 경험적 확증 등이 인간 지식의 기초라는 생각에 동조함
- 마음의 연구가 좀더 경험적으로 이루어져야 한다는 원칙 하에 자신의 기억을 조사함
- 그러나 기억은 모호한 개념이므로, 잘 정의되고 관측 가능한 몇 가지 절차들로 대체하여 두 가지 연구 주제를 정함. “우리는 얼마나 빨리 배우고 빨리 잊어버리는가?”
- 그러나 인간의 기억은 이러한 실험에 적합하지 않음. 모든 개별 정보는 개념들의 네트워크에 내장되어 있어서 우리는 그것들을 사실, 생각, 감정, 장소, 시간, 사건 등과 연결시킴
- 특정 사실을 기억하는 우리의 능력은 다른 요소로부터 떼어내서 연구하기가 매우 어려움
- 이 문제를 해결하기 위해 에빙하우스는 대부분의 연상이 기억하고자 하는 것의 소리와 의미와 상관 있

- 다는 사실을 깨달음
- 원치 않는 연상들을 최소화하기 위해 2,300개의 의미 없는 음절들로 구성된 인조 어휘집을 고안[188]
- 학습을 측정하기 위해 자신의 어휘집에서 무작위로 뽑아서 각각의 음절을 실수 없이 낭독함
- 하나의 목록을 학습하고 정해진 시간을 기다린 다음 자신이 얼마나 많이 기억하고 있는지를 확인함
- 목록의 길이, 하루 중의 시간대, 암기하는 데 들인 시간, 특정 음절들이 목록에 배치된 위치, 암기를 반복하는 사이의 시간적 간격 등 다양한 변수들을 조합하면서 체계적으로 탐구
- 암기할 대상의 수와 시간 사이의 관계를 ‘학습곡선learning curve’으로 분석함
- 에빙하우스는 망각에 대해서도 중요한 발견을 했는데, 20분만 지나도 목록의 절반을 잊어버린다는 것과 이후에는 속도가 느려져 한 달이 지난 뒤에도 목록의 5분의 1을 기억함
- 에빙하우스가 발견한 망각과 시간의 관계는 ‘망각곡선forgetting curve’라고 불림[189]

### 집단기억과 집단망각에 관한 연구

- 어떤 일들은 잊어버리기가 매우 어렵다. 비행기 두 대가 뉴욕의 무역센터를 향해 돌진한 사건이 벌어지고 10년여 년이 흘렀지만 그날의 기억은 여전히 미국인들을 괴롭힌다.
- 9-11의 아침으로부터 거의 60년 전에 미국인들은 자신들의 영토가 최초로 공격받았다는 소식에 잠을 깼. 1941년 12월 7일 진주만 공격. 이 공격으로 미국인 2,400명이 죽고 1,000명 이상이 다침. 이 충격적인 뉴스는 미국을 2차 세계대전 속으로 끌어당김
- 그러나 반세기 이상이 지난 지금, 진주만 공격은 일상의 대화에서 자주 등장하지 않음. 지금은 상상하기 힘들지만 9-11도 이와 같은 과정을 밟을 것
- 망각은 너무나 기이하고, 다른 생각들에 매우 의존하기 때문에 실험이 어려움
  - 1) 미국이 1차 세계대전에 참가하게 된 계기가 된 원양 여객선 루시타니아호 침몰 사건 이후 수십 년이 흐르면서 이 사건은 예상보다 빠르게 잊혀짐. 2차 세계대전 직전에 잠시 되살아남. 제2의 세계대전에 대한 우려가 첫 번째 세계대전 즈음에 있었던 사건들을 전면에 부상시킴. 이러한 ‘연상을 통한 기억’ 효과는 까다로운 문제. 설명하기도 예측하기도 불가능함
  - 2) 장시간에 걸쳐 변화하면서 사람들이 똑같은 사건을 다른 방식으로 기억하고 다른 표현을 쓰게 됨. 1차 세계대전은 처음에 ‘대전쟁’이라 불림. 2차 세계대전이 발발하면서 ‘대전쟁’이라는 용어는 재빨리 사라지고 1차 세계대전이라는 용어로 대체됨. 중요한 것은 사람들이 대전쟁에 대해 생각하는 것을 멈추지 않았다는 것. 여전히 집단기억 속에 깊숙이 박혀 있었음. 사람들은 그 전쟁을 두 차례의 물리적 충돌이라는 좀 더 다른 맥락에서 다르게 생각했고, 이에 따라 다른 언어를 사용함. 이런 종류의 효과는 설명하기도 예측하기도 불가능함[192]

- 망각을 측정하려면 에빙하우스를 모방해 세심하게 선택된 어휘집을 사용함으로써 모든 연상효과를 최소화해야 함. 이를 위해 1816, 1952와 같이 연도들에 상응하는 숫자들만 사용해서 집단기억을 조사함
- 사람들이 어느 연도를 얼마나 자주 언급하는지 살펴보면 우리는 그해에 일어난 사건들이 그들의 마음 속에 어떤 의미로 존재하는지 예측할 수 있음

### 망각곡선

- 망각 과정은 두 개의 체제regime로 구성된 것으로 보임
- 특정 연도에 대한 관심은 첫 몇 십 년 동안 급격히 떨어지다가 그 후로는 훨씬 천천히 떨어짐
- 집단기억과 개인기억 사이에는 충격적인 유사성이 있음

- 사회 역시 단기기억과 장기기억을 가지고 있음
- 1987년의 빈도는 1896년에 절반으로 줄어들었음. 반면 1973년은 겨우 10년 뒤인 1983년에 절반으로 떨어짐. 1973년은 좀 더 빠른 감소를 보여줌
- 세월이 흐를수록 집단망각의 반감기는 점점 더 짧아짐. 이런 관측은 과거에 대한 우리 사회의 태도가 변하고 있음을 시사함. 우리는 점점 더 빨리 과거의 사건에 흥미를 잃고 있음

### 오래된 것들의 퇴장, 새로운 것들의 입장

- 집단의식은 단순히 잊어버리는 것 이상을 해냄. 따라서 집단기억을 이해하려면 동전의 다른 면을 조사할 필요가 있음
- 정보화 시대는 정보가 짹짹 놀랄 만한 속도로 한 사람 혹은 한 장소에서 다른 사람이나 다른 장소로 전해짐. 그러나 우리는 지난 수세기 동안 날것의 정보가, 우리가 더 이상 진가를 완벽하게 알아보지 못하는 잠재력 있는 매커니즘을 이용해 얼마나 빨리 이동할 수 있었는지를 파악하는 시야를 잃어버림[197]

### 사회는 점점 더 빨리 배우고 있다

- 사회는 점점 더 빨리 배우고 있음. 왜 그런가? 집단적 망각과 마찬가지로, 우리는 왜 그런지를 정확히 알지 못함. 그러나 상상해보면 그 잠재적 결과들은 대단히 매혹적임
- 집단학습 곡선이 계속해서 짧아진다면 어떤 일이 벌어질지에 대한 가장 흥미로운 예측은 물리학자 스타니스와프 울람과 존 폰 노이만 사이의 대화에서 등장함. 울람은 수소폭탄 발명자이고, 노이만은 유명한 수학자인 동시에 물리학자, 게임이론가, 컴퓨터과학의 창시자
- 노이만은 “기술의 진보와 인간 삶의 방식의 변화는 인류 역사의 순수한 특이점singularity에 접근하고 있다”고 생각함
- 미래학자 레이 커즈와일은 컴퓨터 칩들의 성능이 향상되는 속도가 2045년이 되면 평범한 컴퓨터 한 대가 인류 전체의 뇌를 합한 것보다 더 뛰어난 연산력을 가질 것이라고 시사함. 우리의 생각을 디스크에 다운로드하는 것이 가능해지고, 기계 속에서 영원히 살 수 있을 것이라고 예측함. 기술적 특이점.
- 하지만 커즈와일이 옳은가? 우리는 정말 모른다. 엔그램들은 과거에 대해서만 말해준다. 아, 그들은 미래를 예측하지 못한다. 아직까지는. [209]

### 7장. 유토피아, 디스토피아, 데이터토피아

어느 누구도 이것을 정확히 뭐라고 불러야 할지 모른다. 그리고 이곳이 어디로 가고 는지 아는 사람도 없다. 그러나 한 가지는 확실하다. 과학과 인문학이 다시 한 번 같은 목표를 향해 가고 다는 것이다. 갈릴레오가 17세기에 우리 세계에 대한 이해를 바꿔놓았듯이, 21세기에 이 두 개의 렌즈는 서로 등을 맞댄 채 갈릴레오가 했던 것과 똑같은 일을 해낼 것이다.[220]

- 이 책에서 디지털화한 역사적 기록들이 전과 달리 어떻게 우리 인류 집단을 수량화할 수 있는지 살펴봄
- 이 책에서 보여준 단순한 도표들은 거대한 빙산의 일각에 불과함
- 다가오는 시대에는 디지털화한 개인적·역사적 기록들이 우리 자신과 우리 주변의 세계에 대해 생각하는 방식을 완전히 바꿀 것임
- 마지막 물음 “빅데이터는 약속의 땅으로 판명될까? 아니면 다가오는 몇 년 동안 우리가 내리는 결정들은 우리에게 다시 전염병을 불러올 것인가?” [221]

## 책

- 앤그렘 뷰어 초기에는 그동안 출판된 책의 4퍼센트, 다시 말해 스물다섯 권 가운데 한 권으로 구동됨
- 2012년에는 6퍼센트, 즉 열일곱 가운데 한 권을 포괄하도록 업그레이드 함
- 만약 구글이 디지털화한 3000만권을 모두 포함시킨다면 출판된 모든 책의 20퍼센트를 약간 상회할 것
- 남은 80퍼센트의 오래된 책들, 다소 불편하게도 물리적인 물체로만 존재하는 책들이 남아있음
- 사기업들과 각국 정부는 우리의 집단적 유산을 보존하는 동시에 그것으로부터 이윤을 만들어내려는 열망으로 움직이기 시작함
- 구글은 이 세상에 존재하는 1억 3,000만 권의 책 가운데 3,000만 권을 이미 디지털화 함
- 2020년까지 모두 디지털화할 수 있으리라 전망함
- 책 기록을 포괄하는 범위가 4퍼센트에서 100퍼센트로 스물다섯 배 개선된다면 우리가 문화의 망원경으로 볼 수 있는 관측의 종류는 굉장히 달라질 것
- 맨눈의 30배율인 망원경을 가지고 지구를 우주의 중심이라는 지위에서 쫓아낸 갈릴레오를 떠올려보라

- 책 기록에 대한 연구는 주요한 장애에 직면해 있음
- 심각한 장애물의 하나는 저작권법. 1998년의 ‘저작권 보호기간 연장법’으로 저자의 사후 저작권을 70년으로 연장함. 이 법은 1923년 이후 출판된 거의 모든 책의 온라인 전파를 간단히 금지했고, 디지털 연구자나 디지털 도서관들에 어떠한 대비책도 마련해주지 않음
- ‘인터넷 아카이브’ ‘해시트리스트’ ‘구텐베르크 프로젝트’ 등의 조직들은 책들을 최대한 공개적으로 사용할 수 있도록 분투하고 있음
- 디지털 책들과 디지털 정보가 중요해지면서 물리적인 책의 생존이 몇몇 전선에서 위협받고 있음
- 킨들 전자책 리더 플랫폼이 소개된 후 3년 만에 아마존 킨들 판매량이 인쇄된 책들을 추월함
- 물리적인 책들은 과거 자신들의 요새였던 도서관에서도 위협을 받고 있음
- 온라인 도서관들이 계속해서 큰 발걸음으로 성큼성큼 앞으로 나아가면서, 오프라인의 전통 도서관은 상당히 위축되고 있음. 최근 몇 년 동안 도서관의 60퍼센트가 예산이 동결되거나 축소됨
- 기금이 줄고 공간이 더욱 협소해지면서 도서관이 새로운 책의 공간을 마련하려면 오래된 책들을 치워버려야 하는데, 도난방지를 위해 책에 장착된 추적 장치들 때문에 거저 나눠줄 수 없는 상황
- 추적 장치를 제거하려면 비용이 너무 많이 들어서 주기적으로 비밀리에 책들을 파괴하고 있음
- 대형 도서관은 때로 한 번에 수십만 권을 제거해버림. 우리가 무엇을 잃어버리고 있는지 기록을 남기려는 노력조차 없음
- 간혹 어떤 도서관은 구글이 어떤 책들을 디지털화했는지를 점검하여 어떤 책을 제거할지 결정함. 이런 일은 우리의 주요 문화유산에 대한 전면적 공격을 의미함
- 책의 디지털화는 복잡한 유산을 남길 것

## 신문

- 미국 ‘국립인문학재단’은 한 세기 이상에 걸쳐 있는 600만 페이지의 초기 미국 신문들을 디지털화하기 위해 대규모 재원을 지원함
- 오스트레일리아의 ‘트로브 프로젝트’는 1억 개의 뉴스 기사를 디지털화함
- 구글도 신문사 2,000곳의 기록을 디지털화하고 있음

## 미출간 원고

- 오늘날 수많은 훌륭한 원고들이 손으로 베껴 쓴 필사본으로 존재함

- 하버드 교수 아프산네 나즈마바디는 이란에서 가가호호를 방문해, 그들 가족이 여성의 경험과 관련된 어떤 역사 문서들을 보관하고 있는지 조사함. 나즈마바디는 자신이 발견한 모든 것의 디지털 영상을 제작함. 그 결과는 ‘카자르 이란 여성의 세계Women’s World in Qajar Iran’라는 제목의 아카이브 [ww<sup>1</sup>](#) 음

- 평생기록life logging은 우리가 살아가는 세계의 창이 됨과 동시에 우리 몸의 창이 될 것임  
: ‘나이키플러스 퓨얼밴드’ ‘피트비트’ ‘스캐너두 스카우드’ 등
- 평생기록은 우리 몸의 안과 밖에서 무슨 일이 일어나고 있는지 깜짝 놀랄 정도로 많은 것을 기록하게 해줄 것임
- 그렇다면 모든 경험 가운데 가장 쉽게 사라져버리는 인간의 생각은 어떤가?  
: 지난 10년 사이에 과학자들은 신체가 마비된 사람이 생각의 힘으로 인공 팔을 움직이거나, 컴퓨터 마우스를 움직이게 하는 정신적 명령을 무선으로 전달하는 마음기계mind-machine 인터페이스를 개발  
: 혼수상태에 빠진 것처럼 보이는 사람들과 의사소통하는 데 사용되거나 장난감으로 제작됨
- 우리가 그렇게 하기로 선택하기만 한다면, 우리 삶의 기록들은 우리보다 오래 살아남을 것임
- 우리는 우리 존재의 완벽한 연대기를 아이들이나 사랑하는 사람들에게 남겨줄 수 있음
- 그들은 우리가 거둔 성공과 유감, 우리의 지혜와 어리석음을 보고 배울 수 있음 → 디지털 사후세계
- 만약 당신이 내킨다면, 기업에 당신의 평생기록을 팔거나 과학자, 학자들과 공유할 수도 있음
- 미래 도서관의 위인전 코너는 사람들의 삶에 대한 이야기만 있지 않고 완벽한 방송도 함께 있을 것 [240]

#### 디지털 기록의 양면성

- 범죄수사 [240]
- 사이버 불링 [241]

#### 데이터는 권력이다

- 우리는 바위에 데이터를 새기는 시절에서 멀리 떠나옴
- 머지않아 우리가 경험하는 것의 대부분이 추적하기가 너무 쉬워져, 우리 대다수는 특별히 손쓰지 않아도 모든 것이 그대로 기록되는 일이 훨씬 간단해졌다는 사실을 알게 될 것
- 요컨대, 정보를 보존하는 일은 기술적 난제에서 도덕적 딜레마로 변하고 있음. 이 딜레마는 몇 가지 이슈를 제기함
  - 1) 기록되지 않아야 하는 것은 어떤 것들인가?
  - 2) 만약 기록이 있다면 그것에 접근할 권리는 누구에게 있는가?
- 우리를 시대착오적이라고 부를지도 모르겠지만, 인생기록이 가능해지는 것이 명백한 만큼 공적 인생기록이 아주 위험한 개념이라는 사실 또한 명백함
- 마케터들은 성가신 광고물을 우리에게 쏟아붓기 위해 그것들을 사용할 것
- 마케터들과 글로벌 기업들이 아무런 제약 없이 인생기록에 접근할 경우, 상황이 얼마나 불쾌하게 진행될지 상상할 수 있음
- 반면 기업의 침해는 우리가 우려하는 최악이 아닐 수도 있음. 정부가 모든 시민을 한시도 쉬지 않고 추적하는 데 인생기록을 사용할 수도 있음. 이미 구글이나 페이스북 등은 국가 안보가 위태로울 경우, 그들의 기록을 연방정부에 공개함
- 어떤 경우에는 기업이 원하든 그렇지 않은 정부가 기록을 취득할 수 있음
- 정당한 공공의 이해와 빅브라더 사이의 선은 어디쯤인가? 그 선은 있어야만 한다. 정부가 어떤 사람의 인생기록을 소환할 수 있는 세계에서 저항은 무용지물이다.