

백과사전적 아카이브 Encyve: Encyclopedic Archive

디지털 큐레이션 / 인문지식 빅데이터 / 인공지능과 인문지식

金 炫

한국학중앙연구원 한국학대학원 인문정보학교실

한국학중앙연구원 디지털 인문학 연구소

hyeon@aks.ac.kr



이 저작물(PPT)의 인용 표시 방법:

김현, "백과사전적 아카이브", 제12기 사전편찬교실 교육 자료,
고려대학교 민족문화연구원 사전학센터, 2019. 1. 8.



1. 백과사전과 아카이브의 만남

2. 데이터 기반 인문지식 백과사전

3. 디지털 큐레이션과 인문지식

4. 인공지능과 인문지식

5. 백과사전 편찬의 미래 과제

❖ 백과사전, 아직도 유용한가?

- 오늘날 세계의 어느 유명 출판사도 더 이상 종이로 된 백과사전을 간행하지 않는다. 과거에 백과사전이 누렸던 지위를 이제는 인터넷이 차지하고 있다. 그렇다고 해서, 지난날 인문학자들에게 가장 큰 규모의 지식 사업으로 여겨졌던 백과사전 편찬이 디지털 세계에서 사라지게 되는 것은 아니다.
- 아카이브의 실물 정보가 그 활용성 증대를 위해 확장된 지식으로 나아가려 하고 있고, 학계의 전문적인 학술연구가 대중의 지적 호기심에 닿을 수 있는 실마리를 찾으려 하고 있다. 한층 넓어진 지식 유통의 세계에서는 수요와 공급이 적정하게 만날 수 있도록 하는 지식의 매개자 역할이 더욱 중요하다.
- 그러나 과거의 백과사전 콘텐츠를 그대로 디지털 미디어에 담아내는 방식으로는 그러한 역할을 기대할 수 없다. 초연결 시대의 백과사전은 소통과 협업의 새로운 패러다임에 안에서, 기존의 조직이나 분야의 벽을 넘어서 서로 통섭하는 새로운 체제로 편성되어야 한다.

❖ 百科事典的 아카이브 (Encyclopedic Archive, Encyve)

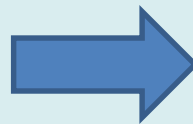
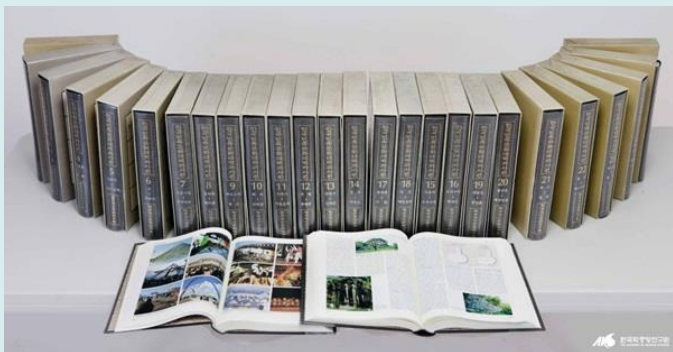
※ 초연결 시대, 백과사전을 중심으로 한 지식 융합 모델의 예시

- 백과사전적 아카이브(Encyclopedic Archive)란 백과사전의 역할을 하는 지식 정보 데이터베이스와 디지털 아카이브를 내용적으로 결합하는 데이터 모델에 대한 구상
- 인문지식의 '원천 자료'이자 그 지식의 진실성을 입증하는 '증거'인 '실물 자료'(기록물, 유물 등) 데이터가 광대한 인문지식 네트워크의 노드(node)로 존재하는 세계
- '자료'와 '해석', 거기에서 파생된 다양한 부산물이 의미의 연결고리를 좇아 서로 이어질 수 있도록 하는 것.

👉 인문지식 시맨틱 웹으로 확장되는 디지털 아카이브

❖ 디지털 시대의 백과사전

- 아날로그 시대에 백과사전은 대중들이 분야별 전문지식의 세계로 들어가는 관문의 역할을 담당.
- 디지털 시대에는 인터넷상에 구현된 월드와이드웹이 종래의 '백과사전'의 역할을 대신.
- 종래의 백과사전은 매체의 제약으로 인해 '개설적인 안내' 기능을 넘어서기 어려웠지만, 디지털 환경에서는 → '보다 전문적인 지식' → '그 지식의 근거가 되는 원천 자료'로의 연계가 가능.
- 시맨틱 웹 기술을 기반으로, 유관한 지식의 조각들이 다양한 방향으로 지식의 문맥을 형성할 수 있게 하는 것이 디지털 시대 백과사전의 미래 모습.



❖ 디지털 시대의 아카이브

- 전통적인 아카이브의 임무는 가치 있는 실물 자료를 수집하고 보존하는 것. 오늘날에는 그 실물 자료의 '활용성'을 증대시키는 일이 중요한 과제로 부상.
 - ✓ **아카이브의 정보화:** 실물을 소장하고 관리하는 일을 보조하는 수단으로 정보기술의 활용을 모색: 자료의 수집, 정리, 전시, 대출 등의 업무 프로세스 자동화
 - ✓ **디지털 아카이브:** 소장하고 있는 '실물'의 관리뿐 아니라 그 실물의 의미와 가치를 알리는 '지식'을 함께 다루어야 하는 과제가 대두.
- 아카이브의 실물 자료가 독립적으로 존재하기보다 세상 사람들의 다양한 관심사에 긴밀하게 연계되어 있음을 밝히는 노력 필요.



Digital Encyve

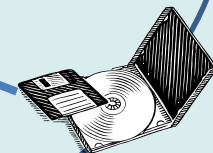
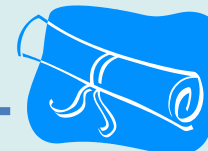
개설적인 안내 → 지식의 문맥을 알리는 다양한 관련 지식 → 그 지식의 근거가 되는 원천 자료로의 연계가 가능



Digital Encyclopedia

Digital Archive

아카이브의 실물자료 하나 하나가 인류, 국가, 지역, 조직의 문화에 관한 지식의 문맥(context) 속에서 하나의 노드(node)로 기능



Environment of Semantic Web

❖ 백과사전적 아카이브의 구현

- 백과사전적 아카이브는 어느 한 곳의 박물관이나 기록관을 그런 형태로 만들자는 제안이 아니라, 백과사전적 지식정보 데이터베이스를 중개자로 삼아 수많은 디지털 아카이브의 실물 정보들이 의미적 관계를 좇아 서로 연결될 수 있게 하자는 것.
- 그 네트워크의 연장은 아주 전문적인 학술 지식에 도달하기도 하고, 매우 대중적인 오락거리 상식과도 접촉.
- 백과사전적 아카이브의 구현은 학계와 아카이브계 양쪽에서 추구되어야 할 과제이지만, 그 실현의 관건이 되는 첫번째 과제는 디지털 지식 네트워크 상에서 운용할 수 있는, 새로운 형태의 '데이터 기반 인문지식 백과사전'을 편찬하는 것.



1. 백과사전과 아카이브의 만남

2. 데이터 기반 인문지식 백과사전

3. 디지털 큐레이션과 인문지식

4. 인공지능과 인문지식

5. 백과사전 편찬의 미래 과제

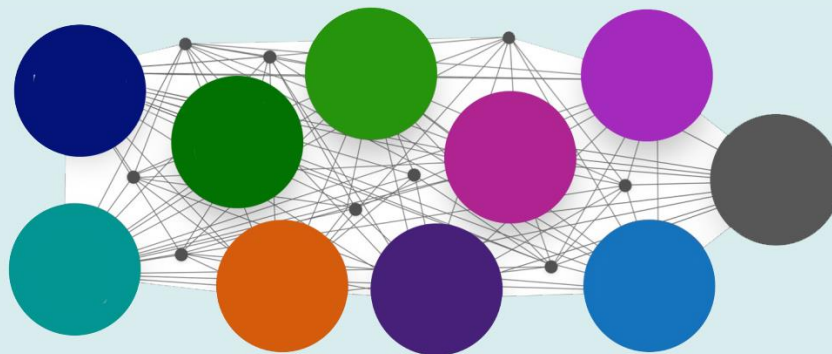
❖ 데이터 기반 인문지식 백과사전, 기존의 백과사전과 무엇이 다른가?

“아날로그의 세계에서 백과사전과 아카이브는 서로 독립적으로 존재했었다. 하지만 이제 그것이 모두 디지털 세계에 있고, 데이터의 형태로 존재한다면, 그 두 가지는 더 이상 별개의 것이어야 할 필요가 없다. 한 번의 클릭으로 아카이브의 실물에 다가갈 수 있는 데이터 시대의 백과사전은 기존의 백과사전과 무엇이 다른가?”

- ① 데이터 네트워크 형태의 백과사전
- ② '지식'과 '자료'를 넘나드는 융합 콘텐츠
- ③ 가상현실에서 지식의 현장을 느끼는 감성적 체험 공간
- ④ 집단 지성의 기여로 확장되는 개방적 데이터

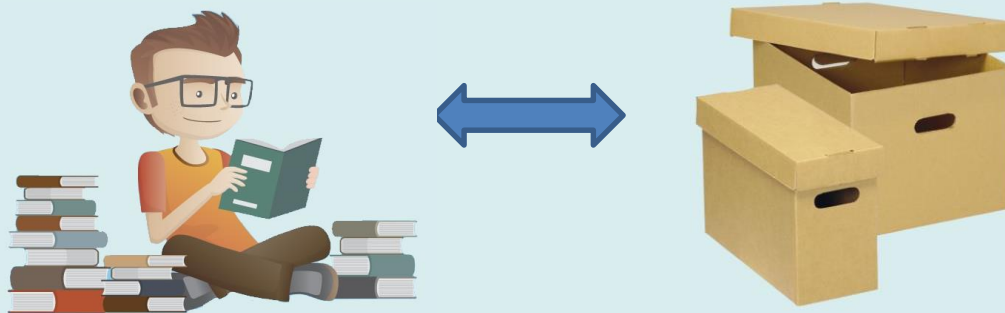
❖ 네트워크 형태의 백과사전

- 다양한 주제의 정보가 독립적인 '항목'으로 나열되어 있는 것이 종래의 백과사전이었다고 한다면, 데이터 기반 백과사전은 항목과 항목, 또는 개별 항목 속의 지식 요소들이 서로 어떠한 의미의 연관이 있는지 그 '관계성'을 보이는 데 주력.
- 데이터 기반 백과사전은 '가나다' 순으로 정리된 목록 형태가 아니라, 방대한 규모의 데이터 네트워크의 형태로 편찬.
- 그 네트워크의 노드 하나 하나가 백과사전의 항목이고, 노드와 노드 사이를 연결하는 연결선은 항목과 항목 사이의 다양한 연관관계를 대변.



❖ '지식'과 '자료'를 넘나드는 융합 콘텐츠

- 디지털 인문학의 주제 가운데 하나는 대학의 강의실에서 다루는 이론적 지식과 아카이브에 소장된 그 지식의 증거 자료 사이의 통섭.
- 아날로그 세계에서는 '연구실'과 '아카이브'가 서로 분리된 영역에 속했지만, 한 번의 클릭으로 그 두 세계를 넘나들 수 있는 디지털 환경에서는 '지식'과 '자료'가 훨씬 가깝게 묶일 수 있다.
- 데이터 기반 백과사전은 서로 관련한 '지식'과 '자료'를 함께 탐구할 수 있는 길을 제공.



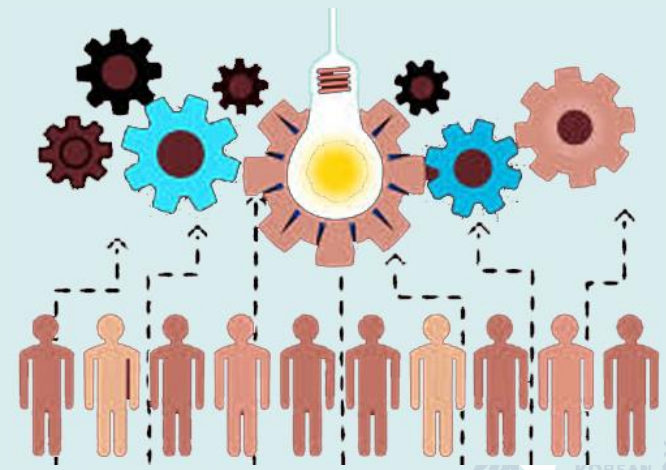
❖ 가상현실에서 지식의 현장을 느끼는 감성적 체험 공간

- 새로운 백과사전의 데이터는 학술적 이론과 아카이브의 유물, 현장의 활동을 하나의 가상 공간에서 융합하고 감성적으로 체험하게 할 수 있게 하는 하이퍼미디어 콘텐츠.
- 역사와 문화의 자취를 담은 지리적 공간을 3차원 가상현실의 속에서 체험하면서, 그 구석구석에서 호기심에 답하는 백과사전적 지식을 열어 볼 수 있는 유저 인터페이스를 제공.



❖ 집단 지성의 기여로 확장되는 개방적 데이터

- 데이터 기반 백과사전의 지식 관계 네트워크는 학자들이 생산한 전문 지식 뿐 아니라, 다양한 층위의 '유용한 지식'에 접근할 수 있는 통로를 제공.
- "위키피디아"처럼 독자가 직접 백과사전 기사(네트워크의 한 노드)를 수정, 증보할 수 있을 뿐 아니라, 노드 사이의 링크 데이터를 추가하여 지식 관계 네트워크를 확장하는 것도 가능.
- 이 방법으로 사회과학이나 산업 분야의 빅데이터에 대한 연결고리가 만들어질 수 있으며, 4차산업혁명을 이끄는 현대 사회의 빅데이터 속에서 고전 인문지식이 유효한 역할을 수행.





1. 백과사전과 아카이브의 만남

2. 데이터 기반 인문지식 백과사전

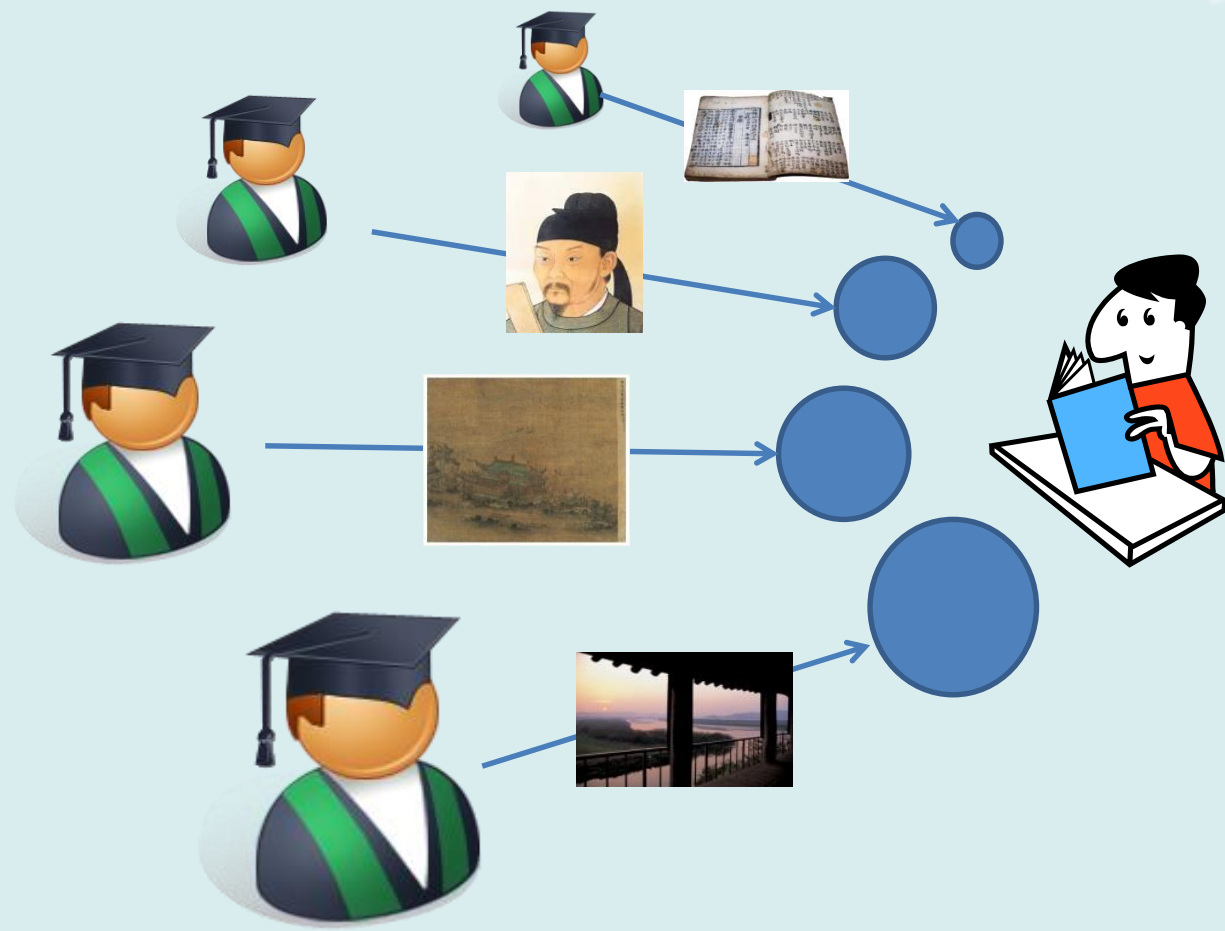
3. 디지털 큐레이션과 인문지식

4. 인공지능과 인문지식

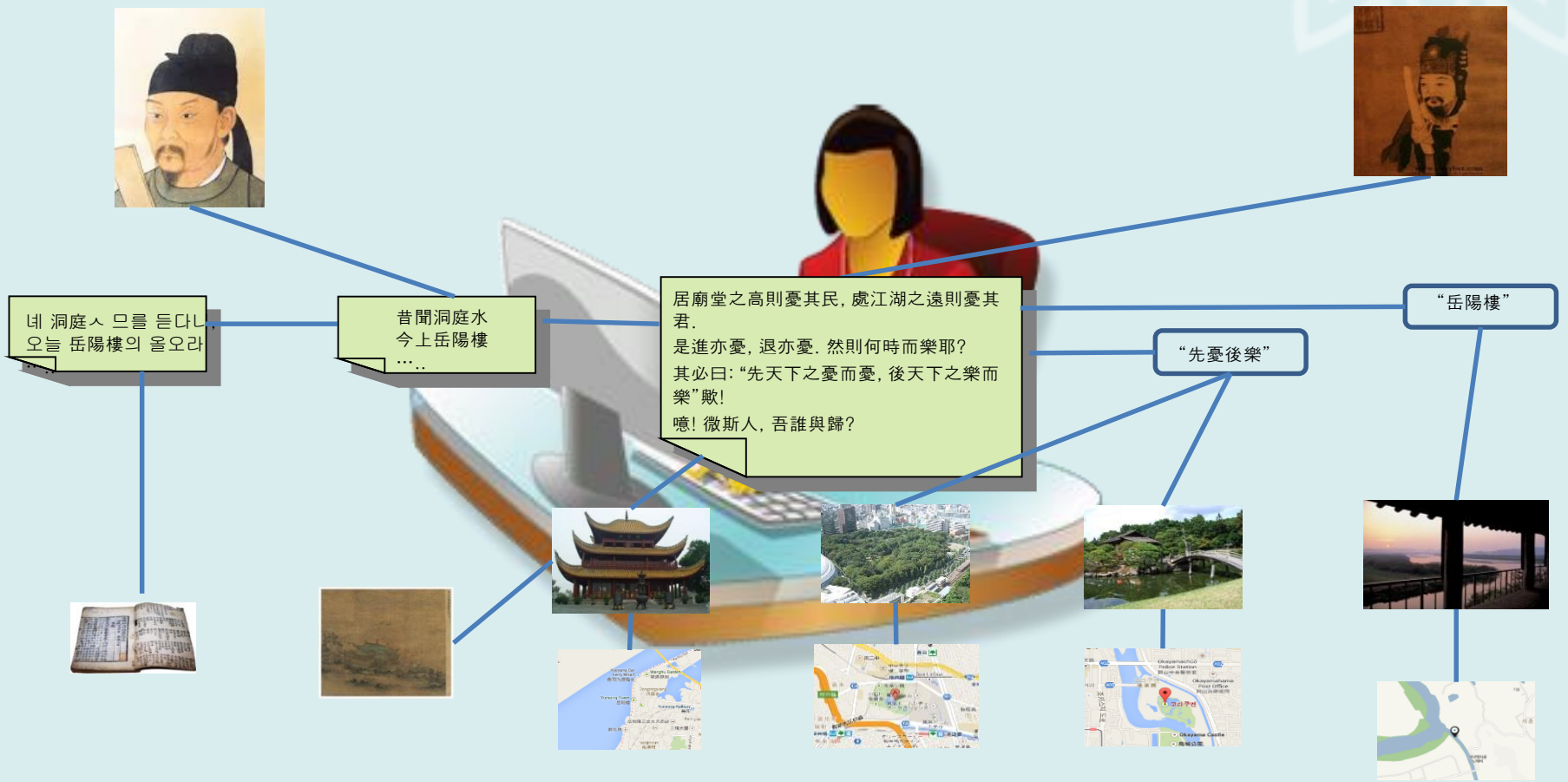
5. 백과사전 편찬의 미래 과제

전통시대의 지식 콘텐츠

❖ 전통시대의 지식 콘텐츠: 각각의 주제(분야, 대상)에 대해 독립적이고 자기완결적인 저작물



❖ 데이터 시대의 지식 콘텐츠: 더 넓은 지식 세계로 통하는 소통의 교점

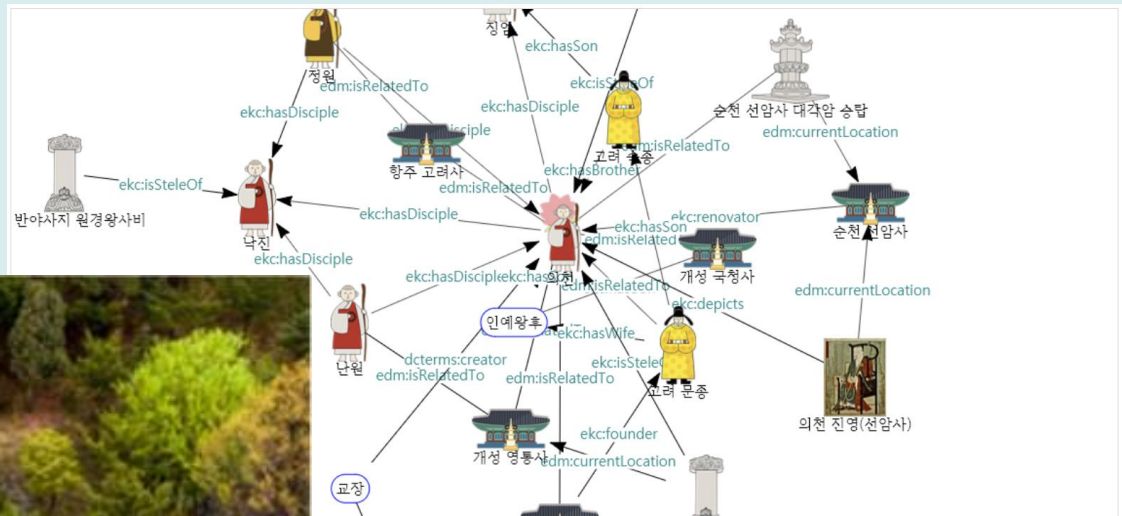


❖ '고립(孤立)'과 '연결(連結)': '죽은 것'과 '살아있는 것'의 차이

- 한 편의 문서, 한 장의 사진, 한 채의 건물, 한 사람의 인물이 각각 고립된 정보로 우리에게 주어진다면, 그것은 단지 과거 사실의 단편적인 조각에 불과.
- 하지만 그 조각들이 서로에 대해 어떤 관계 속에서 만들어지고 존재하게 되었는지를 알게 하는 정보가 주어져서 그 개개의 사실들이 '이야기'로 엮어질 수 있다면, 그것은 바로 오늘날 우리가 이 사회를 살아가는 이야기와도 유사한 콘텐츠.
- 그 이야기의 한 끝에서, 나의 현재의 관심사와 직결되는 연결고리가 발견된다면 그 과거의 이야기는 현재의 이야기의 연장선상에서 새로운 의미와 생명을 얻는 '살아있는 이야기'.

❖ 한국 문화 엔사이브(Enclopedic Archive of Korean Culture)

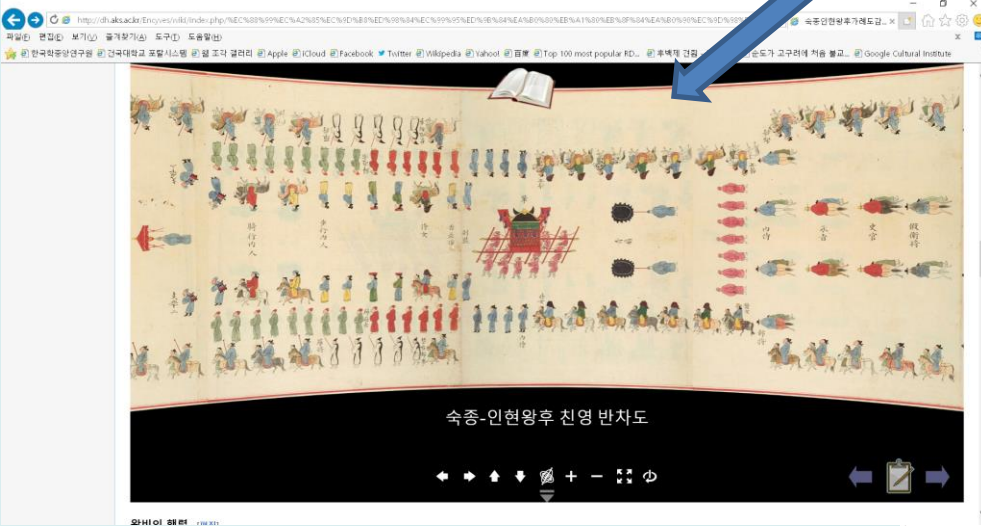
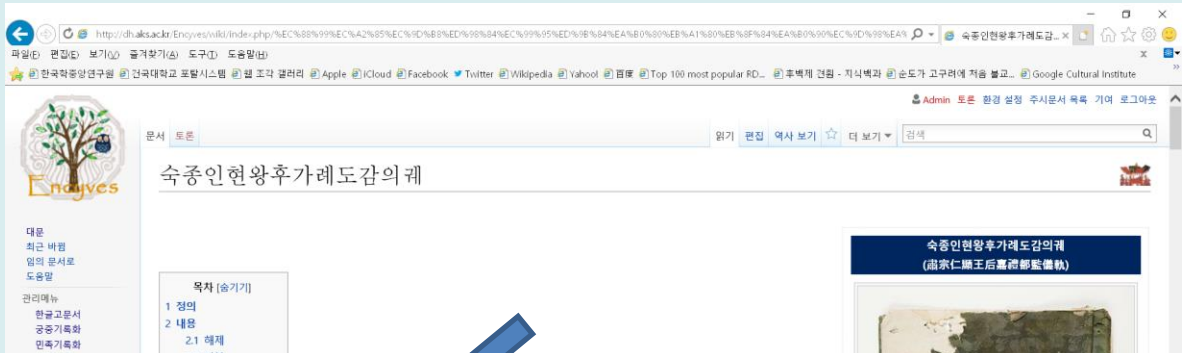
- 명시적인 데이터의 연결을 통해 아카이브의 자료와 백과사전적 지식을 하나로 묶는 것을 추구한 융합 콘텐츠
- 한국의 대표적인 기록문화유산과 관련 지식정보를 디지털 데이터로 편찬하여, 다양한 수준의 한국학 교육에 활용





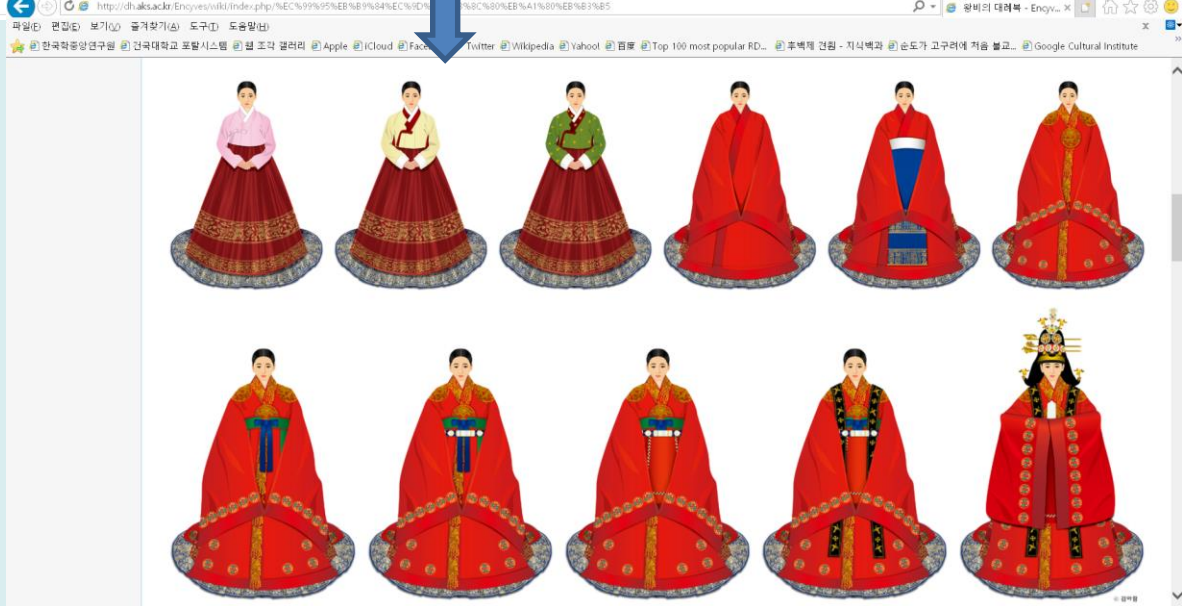
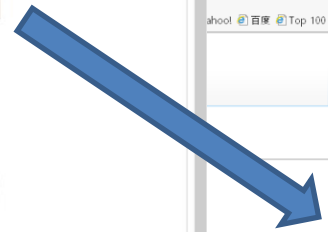
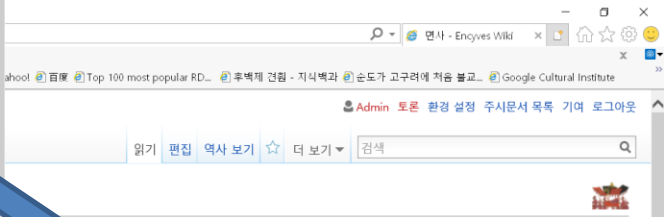
❖ 궁중기록화 가상 전시관

- 대표적인 궁중기록화를 대상으로 삼아 궁중 의례의 내용과 ,행사에 참여한 인물들이 착장하였던 다양한 복식을 디지털 공간에 재현



3. 디지털 큐레이션과 인문지식

디지털 큐레이션 교육 연구 사례



면사 (面紗)

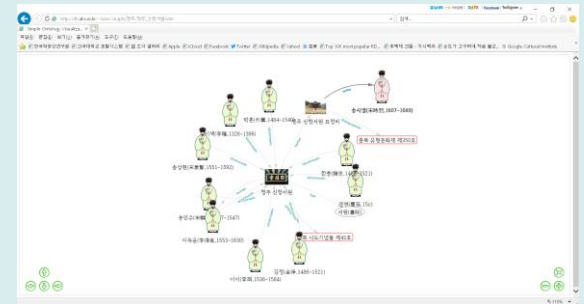
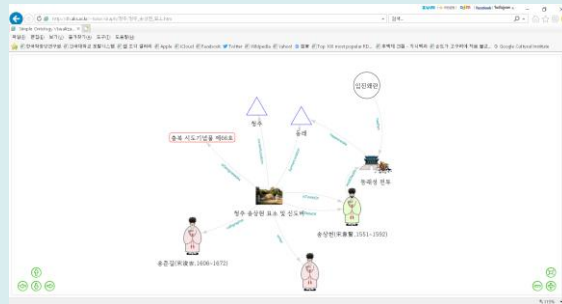


백화여자대학교 전통의상과 (<http://www.baehwa.ac.kr/mbs/kr/>) 소장.

대표명칭: 면사
 한자표기: 面紗
 구분: 관 및 쓰개
 착용신분: 왕비, 왕세자빈, 왕세손빈, 왕녀, 외명부

❖ 문화유산 영문 해설문 편찬 및 시맨틱 데이터베이스 구축

- 개별적인 문화재의 영문 해설문을 편찬하는 과정에서 조사된 관련 지식을 시맨틱 데이터로 기술
- 서로 관련 있는 데이터들이 자동적으로 모여 빅 데이터를 이루게 함으로써 지식의 확장과 새로운 이야기의 발견을 도모



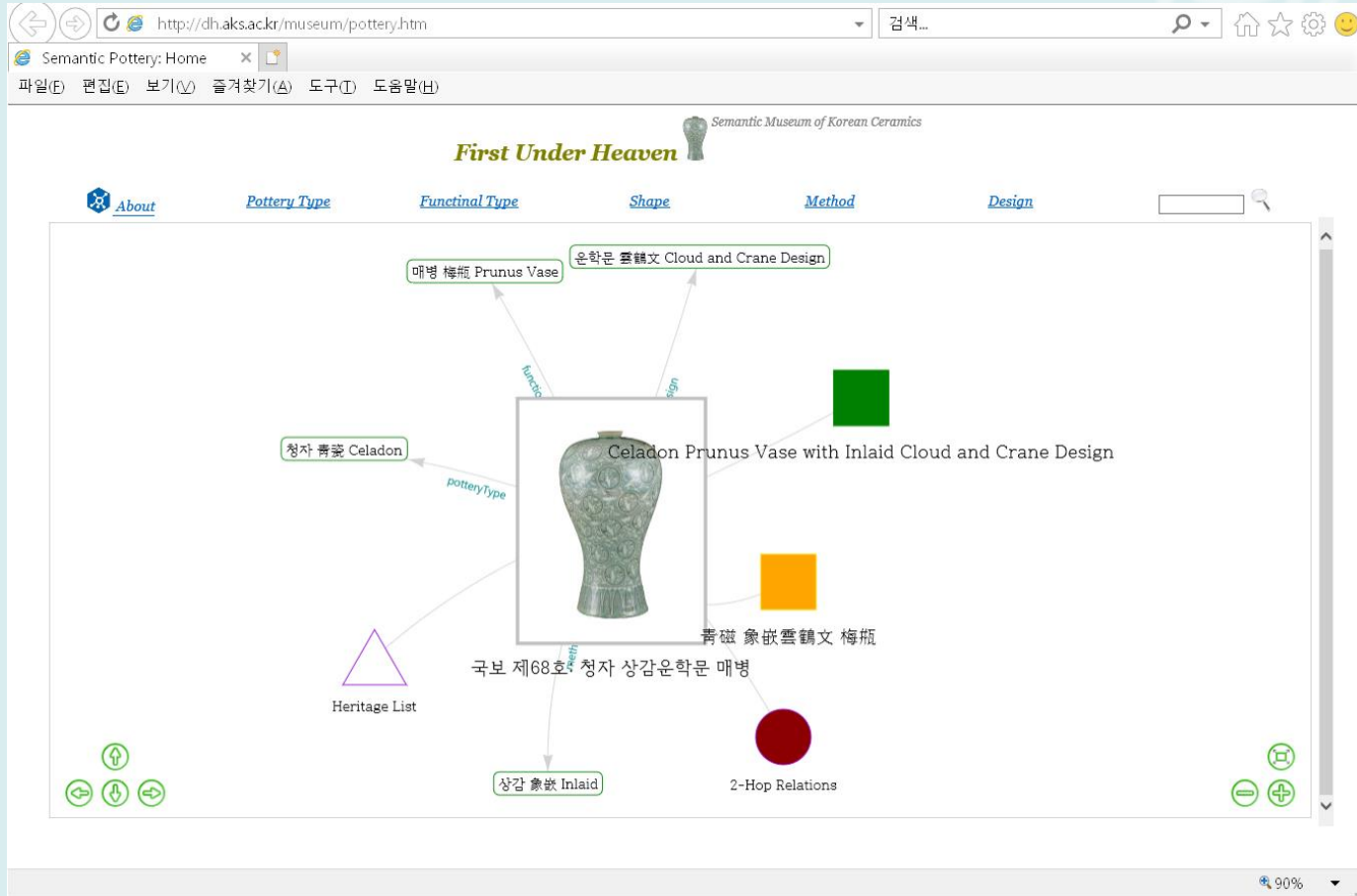
개별 문화유산 데이터

❖ 문화유산 영문 해설문 편찬 및 시맨틱 데이터베이스 구축



시맨틱 네트워크 상에서 자동적으로 확장된 시맨틱 데이터

❖ 문화유산 영문 해설문 편찬 및 시맨틱 데이터베이스 구축



토도자 공예 Semantic Glossary

❖ 인문지식의 디지털 큐레이션

- 인문지식의 디지털 큐레이션은 인간의 언어로 표현되어 온 인간의 지식을 명시적인 데이터로 기술하여, 새로운 소통과 융합의 길을 여는 것
- 인문지식 디지털 큐레이션의 일차적 관심사는 의미있는 '인간의 이야기'를 담은 '스몰 데이터'(Small Data)의 생산과 그 작은 이야기들이 서로 이어져서 '빅 데이터'(Big Data)가 될 수 있게 하는 개방적 소통 모델의 설계.
- 온라인 상의 개방적 협업 공간에서 개개의 '스몰 데이터'(Small Data)는 의미의 연결고리를 좇아 서로 묶이게 되고, 자동적으로 방대한 규모의 인문지식 '빅 데이터'(Big Data)를 형성.
- 이렇게 만들어지는 '빅 데이터'(Big Data)는 인공지능의 학습 데이터가 됨으로써 인문지식의 확장과 새로운 발견에 기여.



1. 백과사전과 아카이브의 만남

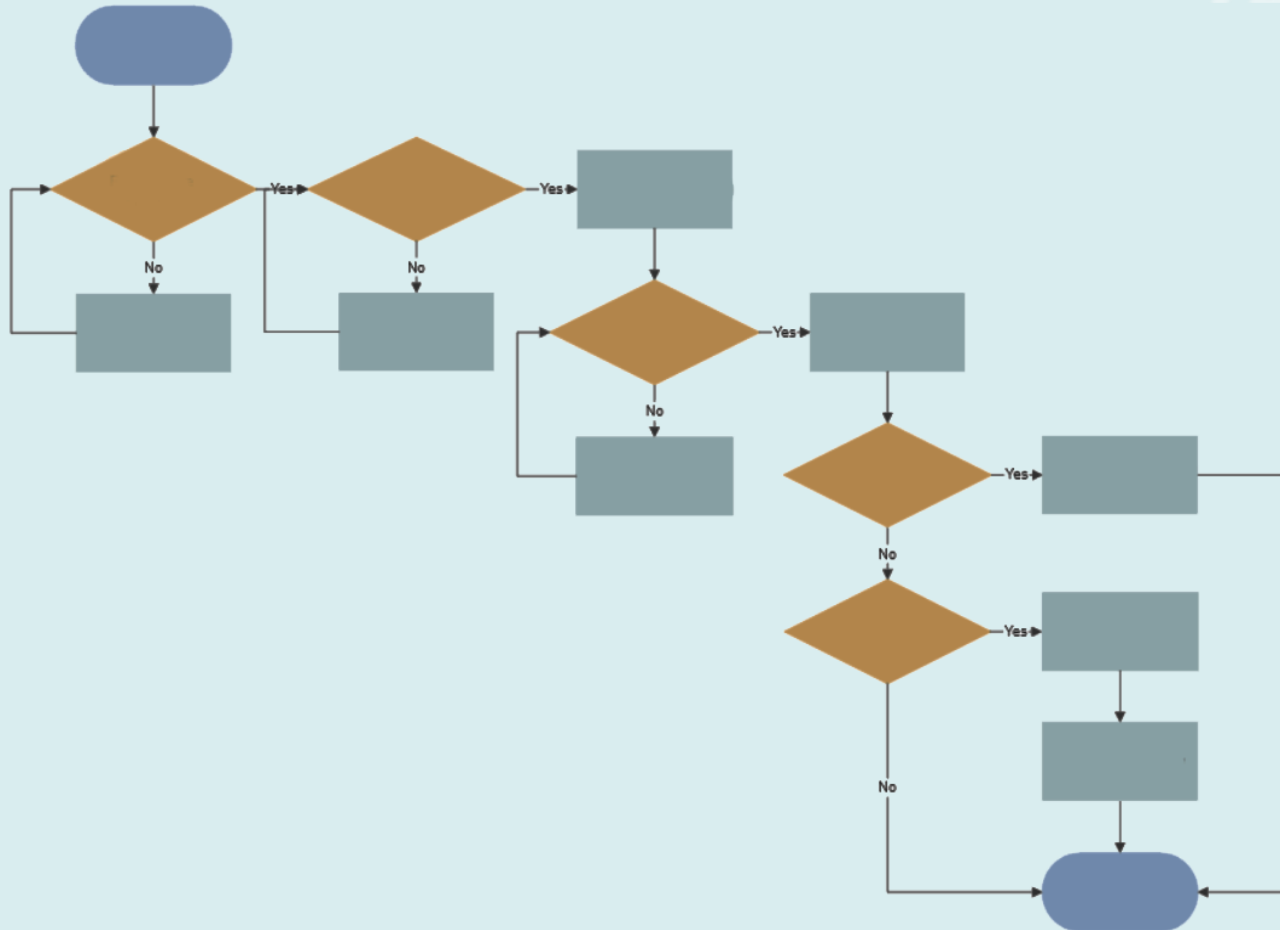
2. 데이터 기반 인문지식 백과사전

3. 디지털 큐레이션과 인문지식

4. 인공지능과 인문지식

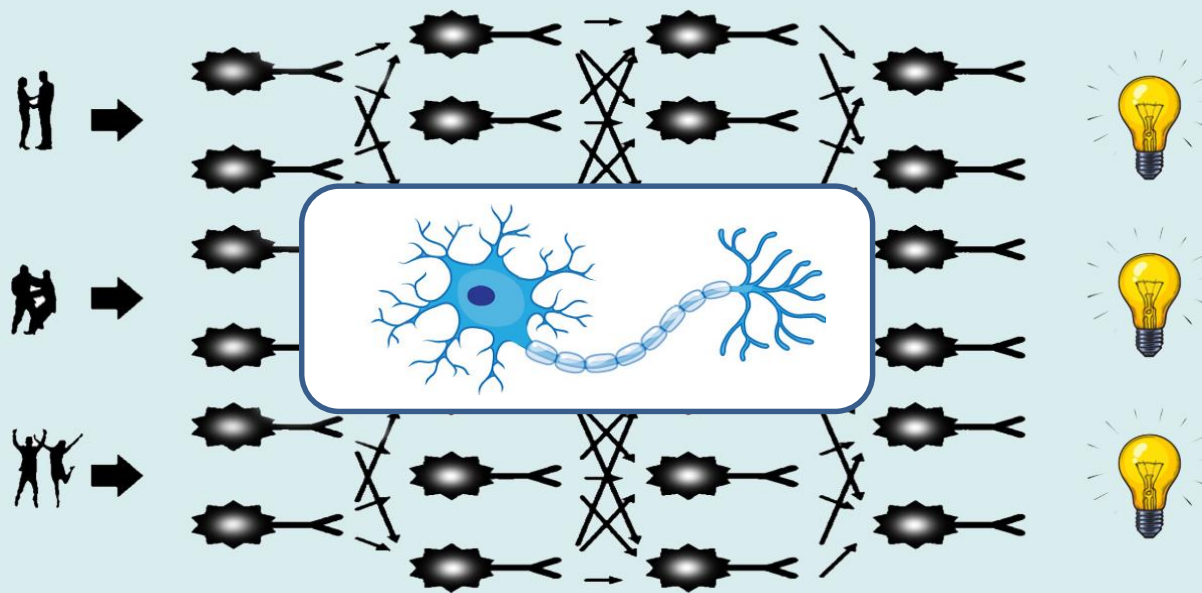
5. 백과사전 편찬의 미래 과제

❖ 과거의 컴퓨터 = 경직된 사고(?)



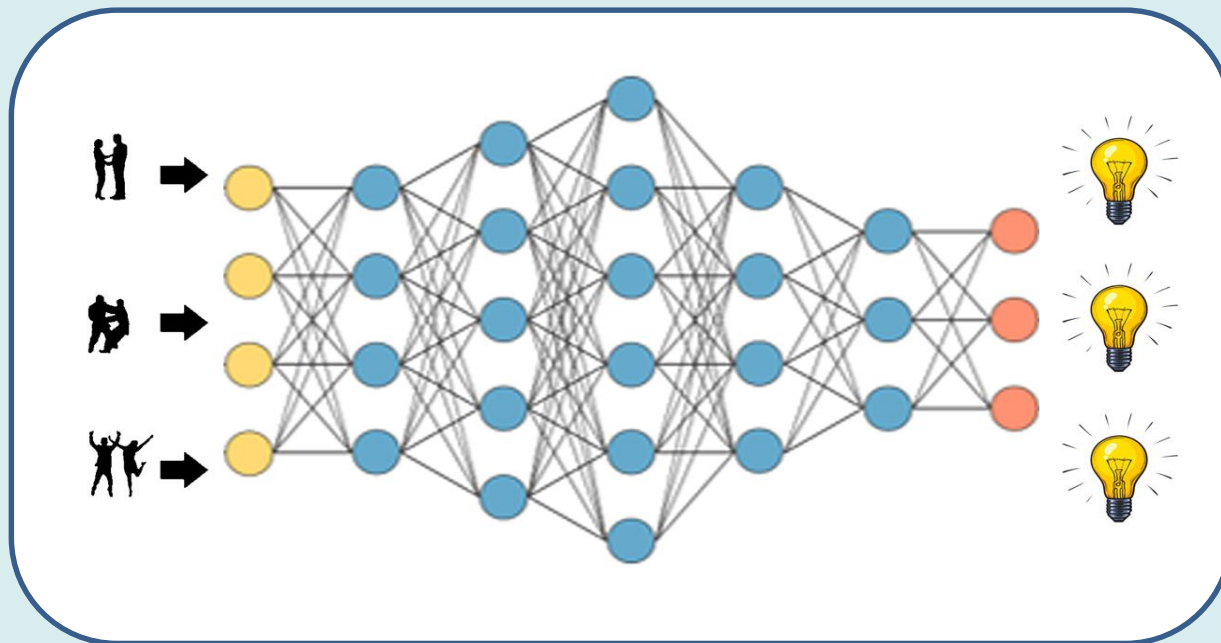
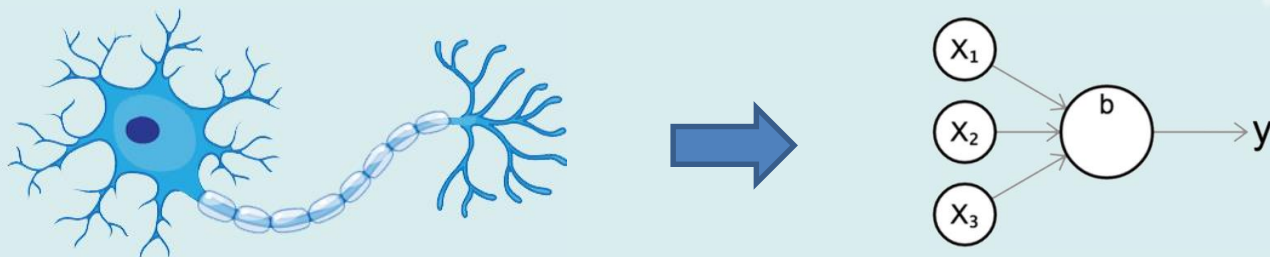
❖ 인간 두뇌의 신경망

- 어마어마한 수의 신경세포(Neuron)가 서로 연결되어 있는 네트워크
- 하나의 뉴런은 다른 여러 뉴런으로부터 여러 가지 입력 신호를 받고 이것을 하나의 출력 신호로 변환하여 다른 뉴런에게 전달
- 이러한 연결의 층이 무수히 많이 존재함으로써 입력 신호에 대한 다각적인 처리와 해석이 가능.
- 반복된 학습을 통하여 가장 유효한 해석의 길을 찾고, 이 학습의 경험을 새로운 입력 신호의 처리에 적용.

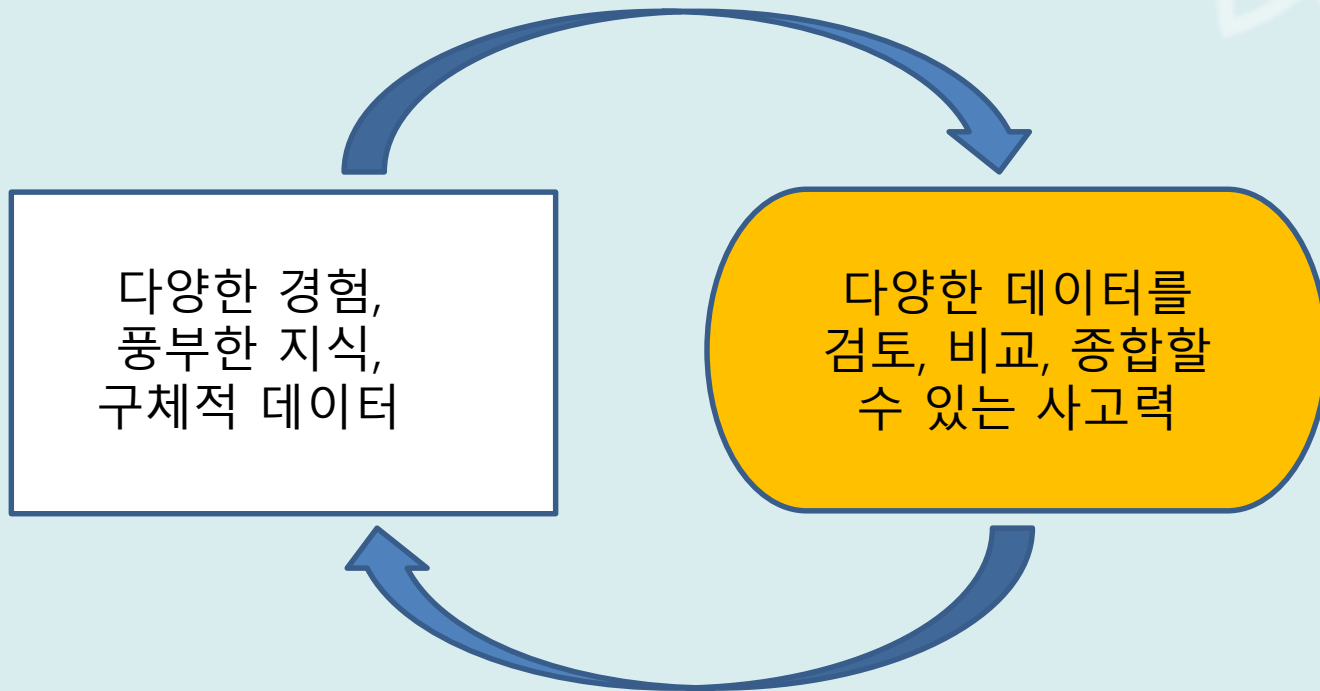


❖ 인공지능

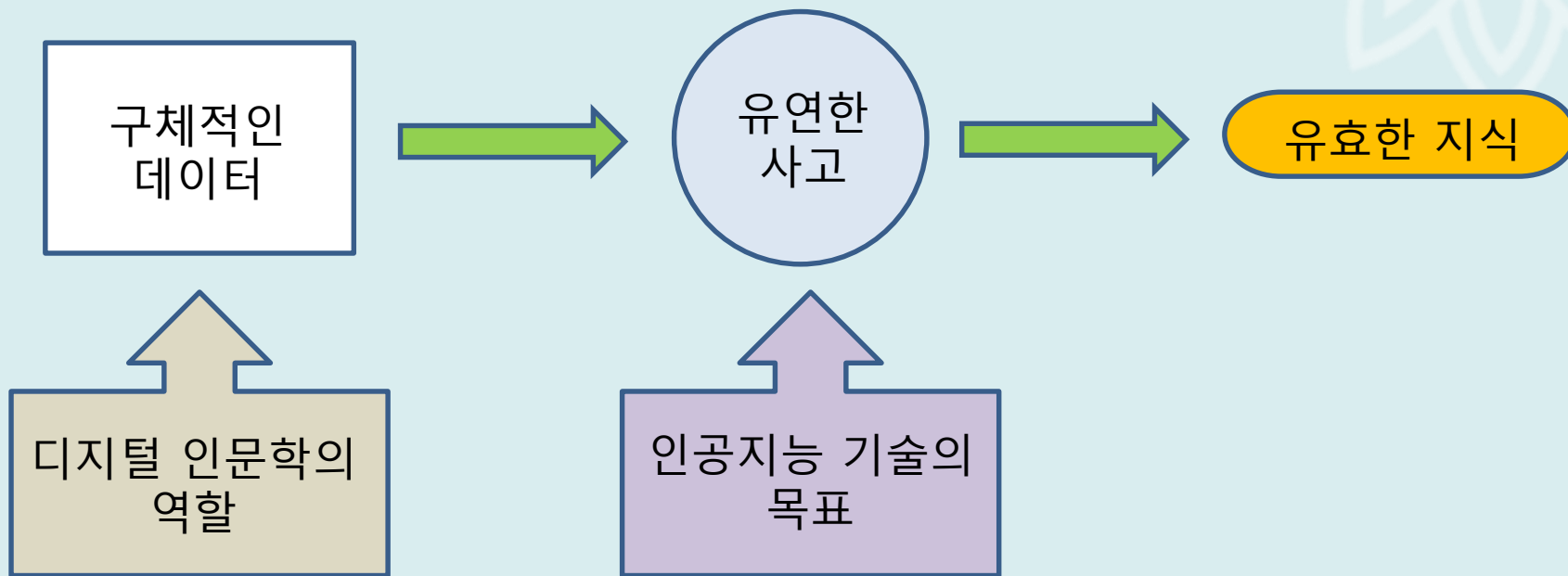
- 컴퓨터의 사고 능력을 신장시키기 위해, 인간 두뇌의 작동 방식과 유사하게 데이터를 처리하게 하는 기술



❖ 유연한 사고의 조건



❖ 인공지능과 인문지식의 만남



❖ 인공지능과 인문지식의 만남에 관한 잘못된 생각

- 인공지능은 구체적인 것에만 매달리고 경직된 사고만 할 수 있다. 인문학의 역할은 자유롭고 창의적인 영감을 불어넣는 것이다.

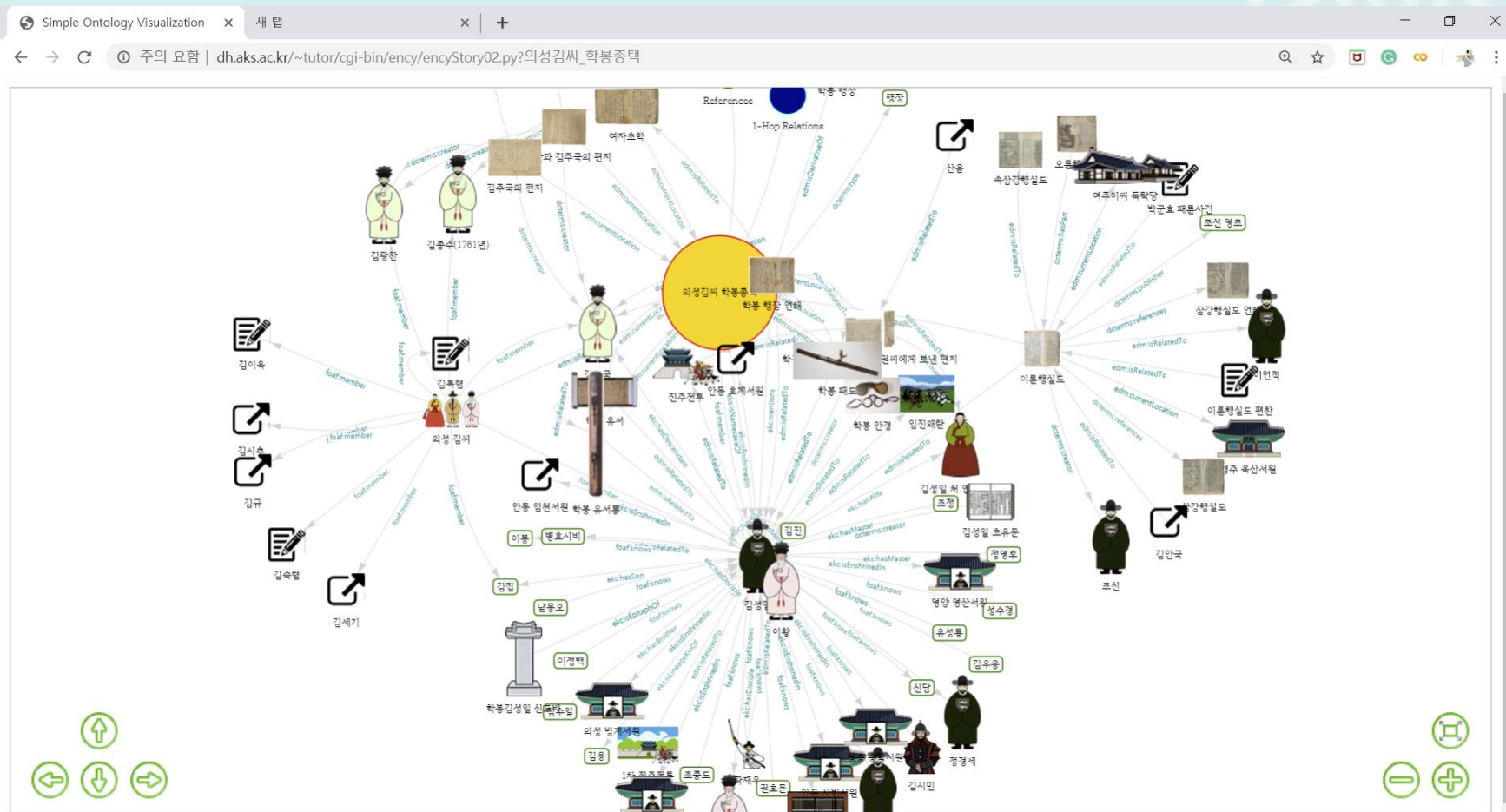
❖ 인공지능에 유효한 도움을 주는 인문학의 역할

- 다양한 데이터를 종합하는 유연한 사고력은 인문학이 아닌, 컴퓨터 기술(인공 신경망, Deep Learning)의 과제.
- 그 사고력이 인간에게도 도움이 되는 유효한 지식 활동이 되기 위해서 필요한 것은 인간에 관한 **풍부하고 구체적인 정보를 담은 명시적인 데이터**.
- 지금으로서는 인간만이 이 데이터를 만들 수 있고, 인공지능을 학습시킬 이 데이터의 우수성이 앞으로 인공지능이 갖게 될 인문적 사고의 유효성을 좌우할 것.

인공지능과 인문지식의 만남

❖ 인문지식 디지털 큐레이션과 인공지능

- 디지털 큐레이션을 통한 인문지식 콘텐츠의 편찬 → 컴퓨터에게 학습시킬 수 있는 명시적인 인문지식 데이터의 생산





1. 백과사전과 아카이브의 만남

2. 데이터 기반 인문지식 백과사전

3. 디지털 큐레이션과 인문지식

4. 인공지능과 인문지식

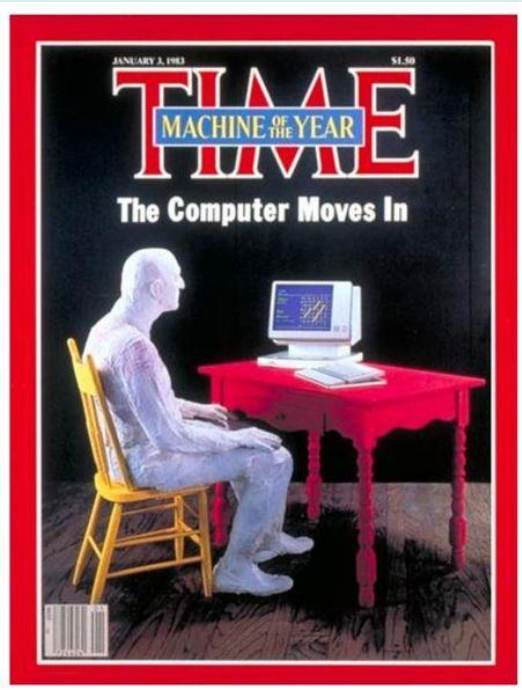
5. 백과사전 편찬의 미래 과제

❖ 디지털 큐레이션: 백과사전의 새로운 편찬 기법

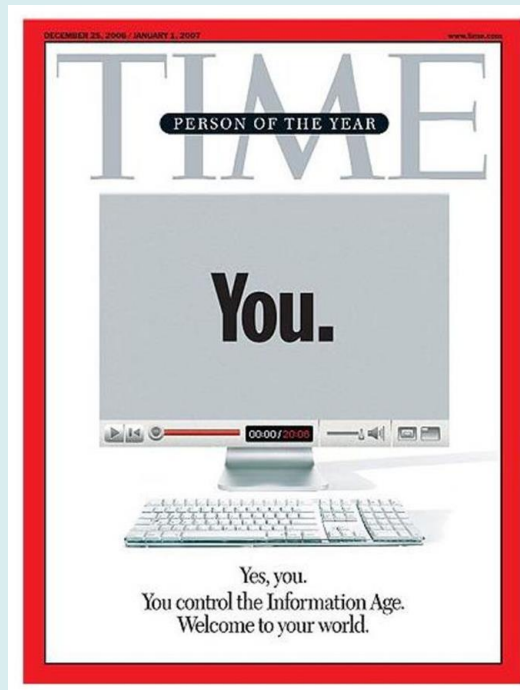
- **디지털 큐레이션**은 디지털 환경에서 인문학을 공부하는 방법이자, 전통적인 인문 지식을 디지털 세계에서 소통할 수 있는 '**데이터**'로 전환하는 방법.
- **인문지식 백과사전의 개별 항목**은 방대한 지식 세계의 네트워크를 만드는 문맥의 노드 → 개별 항목 속의 지식을 디지털 큐레이션의 방법으로 편성하여, **의미있는 '스몰 데이터'(Small Data)**를 생산한다면, → 개개의 '스몰 데이터'는 의미의 연결고리를 쫓아 서로 묶여서 더 큰 데이터가 되고, → 온라인 상의 개방적 협업 공간에서 지속적으로 확장되고 융합하여 방대한 규모의 **인문지식 '빅 데이터'(Big Data)**를 이루게 될 것.
- 이렇게 만들어지는 '빅 데이터'(Big Data)는 **인공지능의 학습 데이터**가 되고, 사회 각분야의 지식 생산에 기여함으로써 한국학 지식의 확장과 새로운 발견에 기여.
- 이러한 노력을 통해 우리의 지식은 전통적인 인문학의 경계를 넘어서 현대 사회의 다양한 영역과 교섭할 것이며, 한국인과 한국사회, 한국이 함께 하는 세계 사회에 대한 이해로서의 한국학의 외연을 더욱 의미있게 확장해 갈 수 있을 것임.

Computer, You, AI

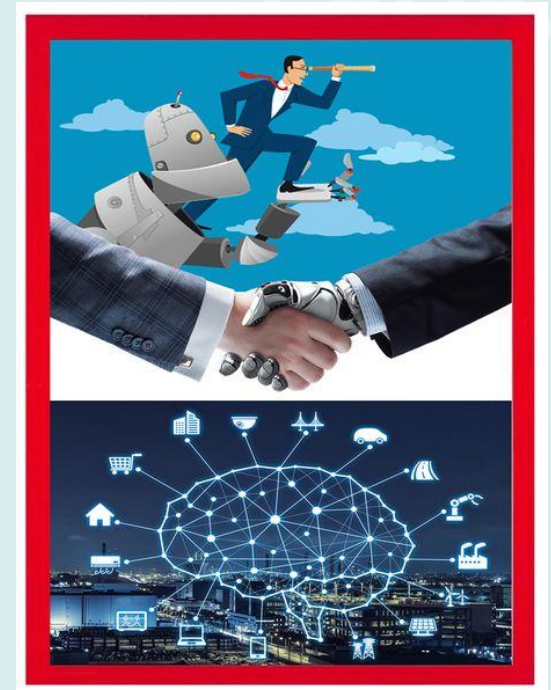
❖ Computer, You, AI



1982



2006



2020~