

제15회 예비-초보 전문가를 위한 한국사교실
2026년 2월 12일 (목)

AI 시대, 디지털 한국사 연구의 개념과 방법

허 수(서울대 역사학부)

강의 개요

- 대상: 역사학 전공 학부·석사 재학/수료생
- 시간: 75분

【강의 목표】

- 예비 연구자로서 **연구와 전공의 의미** 이해
- 한국사 연구의 **패러다임 변화**와 현재 주류 경향 파악
- **디지털 한국사 연구**의 필요성과 방법론적 특성 이해
- 디지털 연구가 인문학·역사학과 괴리되지 않음을 인식

**"디지털 한국사 연구는 선택이 아니라 필수이다.
적극적으로 배우고 익혀야 한다."**

— 오늘 강의의 핵심 메시지

강의 구조

기(起) 연구자로서의 첫걸음 15분

연구란? 한국사란? 전공이란?

승(承) 역사 연구의 심화 15분

연구 설계, 좋은 연구, 연구 경향 변화

전(轉) 디지털 한국사 연구의 특징 30분

1. 왜 디지털인가
2. 디지털 연구방법의 변별성

결(結) 구체적 예시와 마무리 15분

DTM, Neo4j 활용 사례

기(起)

연구자로서의 첫걸음

Q1. 연구란 무엇이고, 연구자란 어떤 사람인가요?

【연구의 정의】

- 알려지지 않은 것을 탐구하여 **새로운 지식을 생산**하는 행위

【학부 vs 대학원】

- 학부: 기존 지식 습득 → "이것이 정답입니다"
- 대학원: **새 지식 생산** → "정말 그런가? 왜 그런가?"

【연구자의 정체성】

- **질문을 던지고**, 자료를 통해 **답을 찾아가는** 사람
- 지식의 소비자 → **지식의 생산자**로 전환

Q2. 한국사는 어떤 분야인가요?

【좁은 방식으로 이해하기】

인문학 — 인간의 삶과 문화, 가치를 탐구



역사학 — 시간 속 인간 사회의 변화 탐구

특징: **사료에 기반한 실증**



한국사 — 한국이라는 시공간 대상

한국어 사료, 한국적 맥락

Q3. '전공을 가진다'는 것은 무엇을 의미하나요?

- ① 그 분야의 **언어와 문법**을 익히는 것
고유한 개념어와 논증 방식 학습
- ② 그 분야의 **통설(정설)**을 파악하는 것
현재 가장 널리 받아들여지는 견해 이해
통설 위에서 자신만의 질문을 던지기
- ③ 선행 연구자들과의 **학문적 대화**에 참여하는 것
앞선 연구자들과 동의·반박하며 학문 발전

승(承)

역사 연구의 심화

Q4. 이론 공부와 사료 분석, 뭐가 먼저인가요?

이론/방법론

- 연구의 **시야와 질문**을 형성
- "이 사료에서 무엇을 봐야 하지?"
- "이 현상을 어떻게 설명하지?"

- ⚠ **사료 없는 이론은 공허**

사료 분석

- **구체적 근거와 답**을 제공
- "이 사료가 이것을 보여준다"
- **실증의 토대**

- ⚠ **이론 없는 사료 분석은 맹목**

연구 설계의 실제

【핵심】 둘 사이를 **오가며** 연구가 발전한다

- 처음 세운 가설 → 사료 분석 과정에서 수정
- 사료에서 발견한 것 → 새로운 이론적 질문으로 연결

【조언】

- 처음부터 완벽한 설계를 하려 하지 말 것
- 읽고, 생각하고, 분석하고, 다시 읽고...
- 이 **반복** 속에서 연구가 형태를 갖추

Q5. 좋은 역사 연구란 어떤 것인가요?

① 통설과의 차별성

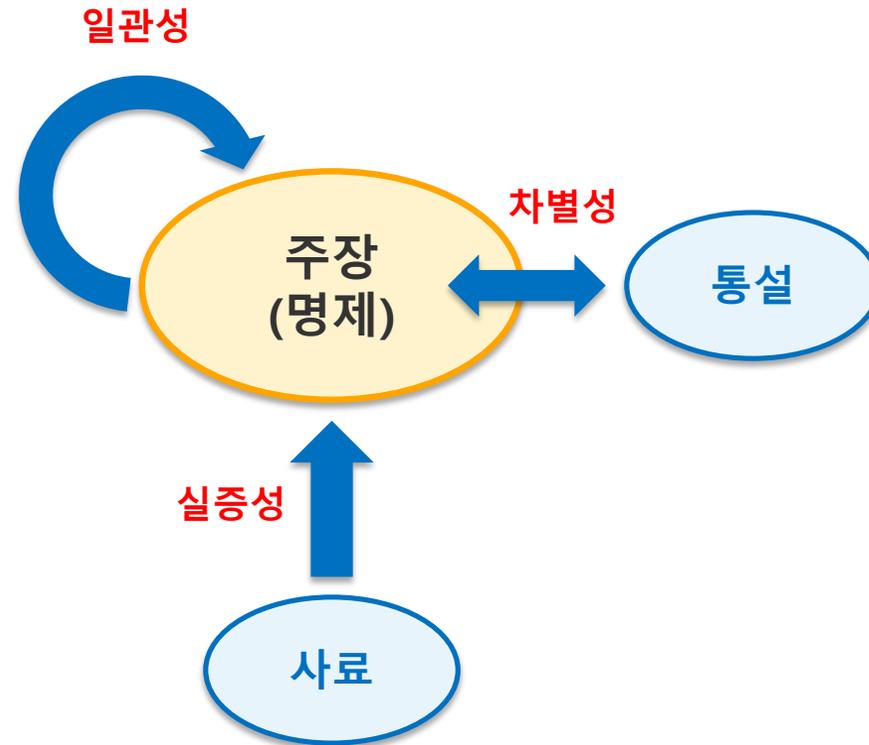
- 기존 연구와 무엇이 다른가
- 새로운 사료 발굴 or 새로운 해석

② 사료에 기반한 실증성

- 주장을 뒷받침할 사료적 근거
- "이 사료가 보여준다" ○

③ 수미일관성

- 연구질문 → 논증 → 답변



Q6. 한국사 연구 경향의 변화

시기	키워드	주요 문제의식
1960년대	독립	식민 지배 극복, 민족 주체성 확립
1980년대	해방	분단 극복, 민중·민족 모순 해결
2000년대	탈식민	식민성 내면화 비판, 포스트콜로니얼
현재	변혁과 번역 ?	지속과 변화, 고유성과 횡적 연관

연구 주제의 확장과 패러다임 전환의 태도

【연구 주제의 확장】

- 과거: 민족, 계급 중심
- 현재: **젠더, 지역, 환경, 일상, 감정** 등 다양화

【현재의 연구 경향】

- **장기적·거시적 시야**
- 지속과 변화, 고유성과 횡적 연관을 함께 포괄

【패러다임 전환 시 태도】

- 이전 세대의 고민과 노력은 중요한 참고
- **계승과 비판의 변증법** — 단절이 아님

전(轉)

디지털 한국사 연구의 특징

전(轉)

1. 왜 디지털인가

Q7. 디지털 한국사 연구가 왜 필요한가요?

외적 요인: 기술 환경의 변화

- **사료의 대규모 디지털화**
 - 한국사데이터베이스
 - 한국역사정보통합시스템
 - 조선왕조실록, 승정원일기 등
- **빅데이터 축적**
- **디지털 분석 도구와 AI** 급속 발전

내적 요인: 연구 질문의 변화

- **거시적·장기적 시야** 요구
 - 장기 지속 분석
 - 사회 구조 변동
 - 네트워크 분석
- 개별 사료 정독만으로는 한계
- **연구 질문 자체가 디지털 방법론** 요구

Q8. 디지털 한국사란 무엇인가요?

【정의】

디지털 기술과 방법론을 활용하여
한국사 연구의 질문에 답하는 연구 분야

【핵심】

- 인문학적·역사학적 질문을 포기하지 않음
- 그 질문에 답하기 위해 디지털 도구를 활용

【강조】

- 별종의 학문이 아니다
- 기존 한국사 연구의 확장이자 심화

'디지털'의 핵심이란?

① 계산 가능성 (Computability)

기계가 처리할 수 있는 형태로 변환
종이 문서 → 텍스트 파일

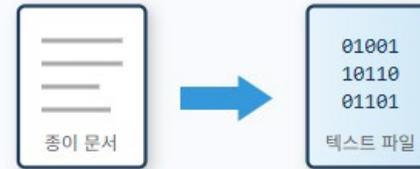
② 이산화 (Discretization)

연속적인 것을 불연속적 단위로 분절
문서 → 문장 → 단어

③ 형식화 (Formalization)

암묵적 지식을 명시적 규칙으로 표현
"관련 있다" → "A는 B의 제자이다"

1 계산 가능성



2 이산화



3 형식화



아날로그 → 디지털 변환의 세 가지 핵심 원리

전(轉)

2. 디지털 연구방법의 변별성

Q9. 디지털 방법론은 무엇이 다른가요?

전통적 방법

- 개별 사료의 정독과 해석



- 서사적 기술

- 연구자의 통찰력과

- 해석 능력 중심

디지털 방법

- 대량 사료의 구조화



- 모델링을 통한 패턴 파악

- 숫자와 규칙에 기반한

- 패턴과 관계 분석

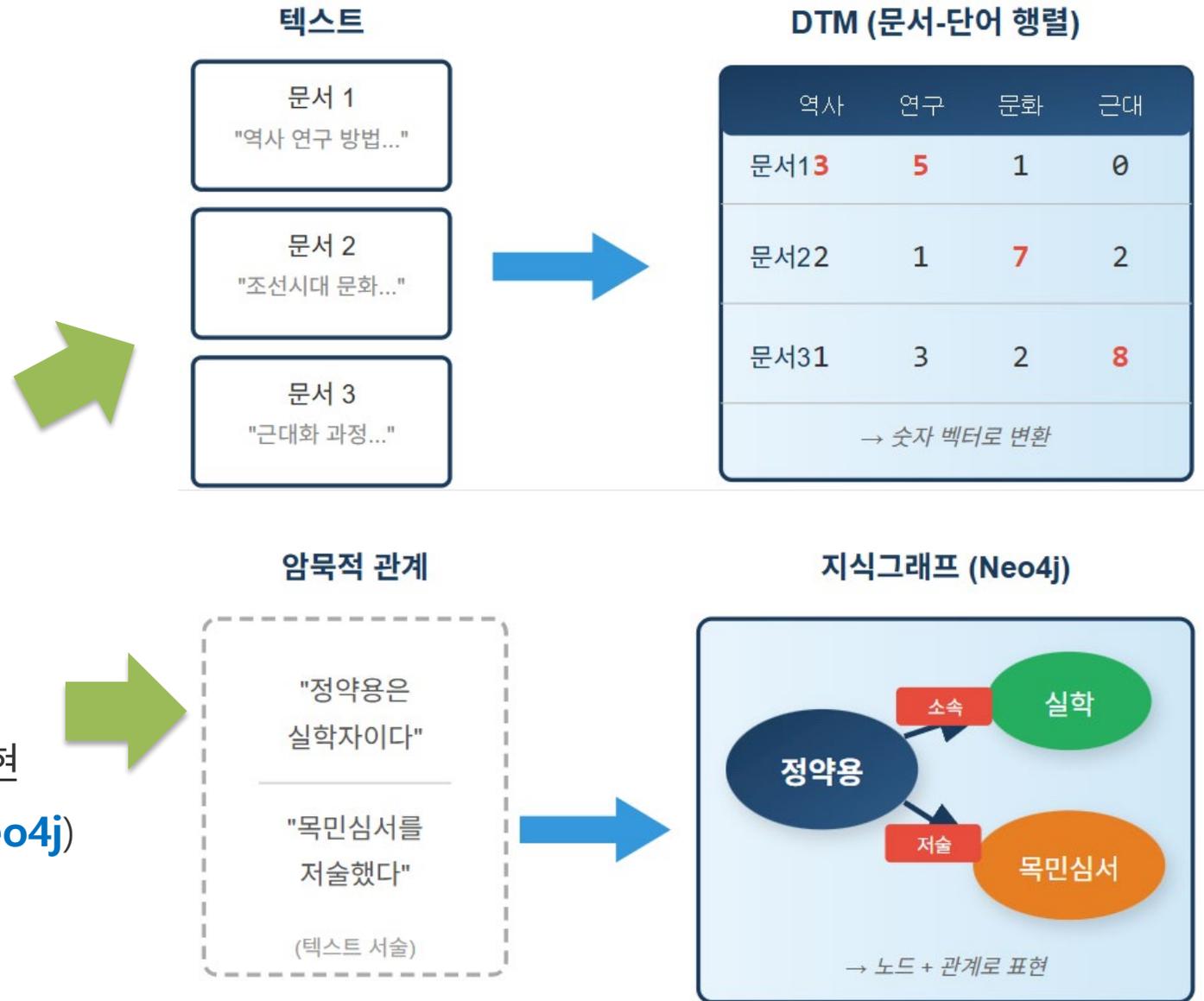
디지털 방법론의 핵심 변별성

숫자와 규칙에 기반한 모델링을 통해
연구 대상의 패턴을 파악

【두 가지 접근법】

① **벡터화(Vectorization)** 기반 접근
텍스트를 숫자 벡터로 변환
대표 기법: **DTM**(문서-단어 행렬)

② **규칙 기반(Rule-based)** 접근
명시적 규칙과 관계를 데이터로 표현
대표 도구: 그래프 데이터베이스(**Neo4j**)



벡터화 접근: DTM (문서-단어 행렬)

【개념】 텍스트를 **숫자 벡터**로 변환하여 계산 가능하게 함

【DTM 구조】

- **행(Row)**: 각 문서 (개별 사료, 논문 등)
- **열(Column)**: 단어 (어휘)
- **값(Value)**: 해당 문서에서 해당 단어의 출현 빈도

【활용】

- 문서 간 **유사도 측정**
- **주제 모델링** (Topic Modeling)
- **시계열 어휘 변화 분석**

규칙 기반 접근: 그래프 데이터베이스 (Neo4j)

【개념】 연구자가 정의한 **명시적 규칙과 관계**를 데이터로 표현

【구성 요소】

- **노드(Node)**: 개체 — 인물, 사건, 기관, 문서 등
- **엣지(Edge)**: 관계 — 저술하다, 소속되다, 참여하다 등

【사이퍼(Cypher) 쿼리】

- 관계를 질의하는 언어
- 예: "이광수와 2단계 이내로 연결된 인물 찾기"

【활용】 인물 네트워크, 기관 관계망, 사건 연쇄 분석

디지털 방법론의 위치

전통적 연구를 대체하는 것이 아니다

【상호 보완적 관계】

- **디지털 방법론**: 거시적 패턴 파악
↓
- **패턴 발견**: "여기에 이런 경향이 있다"
↓
- **전통적 방법**: 해당 지점의 미시적 정독과 해석

→ 새로운 **질문을 발견**하게 해주는 도구

결(結)

구체적 예시와 마무리

예시 1: DTM 개념 이해 — "사과, 바나나" 예제

DTM = Document Term Matrix (**문서단어행렬**)

【세 개의 문서】

문서1: "저는 사과 좋아요"

문서2: "저는 바나나 좋아요"

문서3: "저는 바나나 좋아요 저는 바나나 좋아요"

【DTM 표】

	저는	좋아요	바나나	사과
문서1:	[1,	1,	0,	1]
문서2:	[1,	1,	1,	0]
문서3:	[2,	2,	2,	0]

→ 각 행이 해당 문서의 **특성 벡터**가 됨

두 벡터의 내적(공통점)

$$\text{문서1-2} : 0*1 + 1*0 + 1*1 + 1*1 = 2$$

$$\text{문서1-3} : 0*2 + 1*0 + 1*2 + 1*2 = 4$$

$$\text{문서2-3} : 1*2 + 0*0 + 1*2 + 1*2 = 6$$

벡터의 크기

$$\text{문서1} : \sqrt{0^2 + 1^2 + 1^2 + 1^2} = \sqrt{3}$$

$$\text{문서2} : \sqrt{1^2 + 0^2 + 1^2 + 1^2} = \sqrt{3}$$

$$\text{문서3} : \sqrt{2^2 + 0^2 + 2^2 + 2^2} = \sqrt{12} = 2\sqrt{3}$$

문서 유사도 = (두 벡터의 내적 / 두 벡터의 크기 곱)

$$\text{문서1-2} : (0*1 + 1*0 + 1*1 + 1*1) / (\sqrt{3} * \sqrt{3}) = 2/3 = 0.67$$

$$\text{문서1-3} : (0*2 + 1*0 + 1*2 + 1*2) / (\sqrt{3} * 2\sqrt{3}) = 4/6 = 0.67$$

$$\text{문서2-3} : (1*2 + 0*0 + 1*2 + 1*2) / (\sqrt{3} * 2\sqrt{3}) = 6/6 = 1$$

예시 2: Neo4j 활용 — 메치니코프 사례

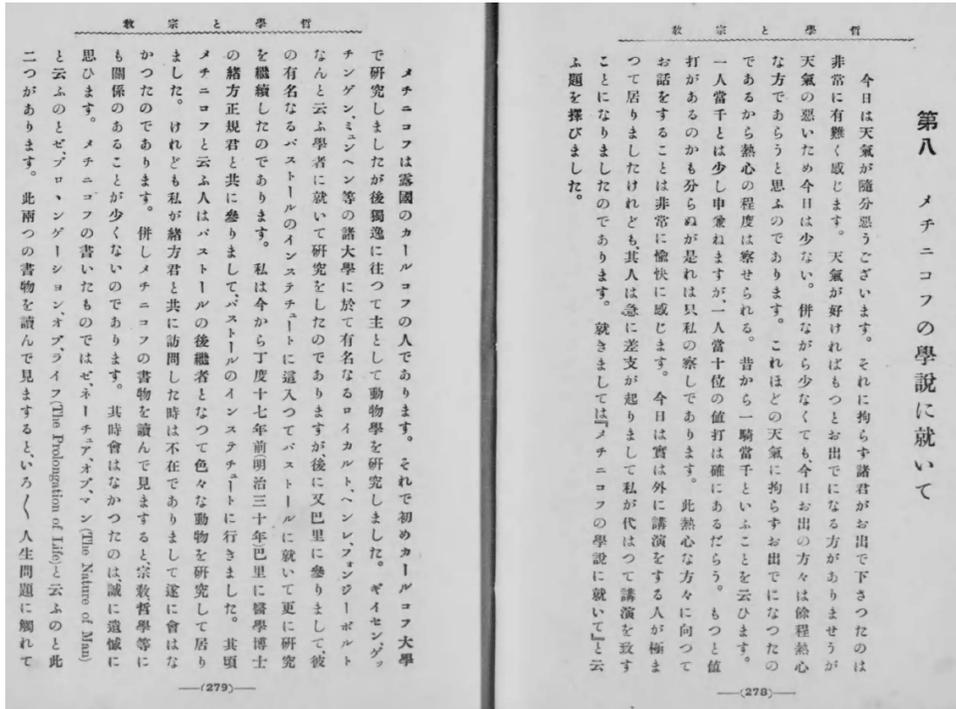
【연구 질문】

식민지 조선의 지식인들은 **외래 지식**을
어떻게 수용했는가?

【분석 대상】

- **원문**: 이노우에 데쓰지로 《철학과 종교》
'메치니코프 학설에 관하여' (**689문장**)
- **번역문**: 이돈화 《개벽》 8호
"과학상으로 본 생노병사" (**175문장**)

분석 대상: 원문과 번역문



【원문】 이노우에 데쓰지로 《철학과 종교》 (1915) 제8장 '메치니코프 학설에 관하여'



【번역문】 이돈화(백두산인) 《개벽》 8호 (1921년 2월) "과학상으로 본 생노병사"

분석 과정

① 문장 단위로 원문-번역문 매핑

② 그래프 데이터베이스 구축

노드: 문장, 어휘, 인물

엣지: 발췌, 번역, 삭제, 포함 등

③ 사이퍼 쿼리로 관계 질의

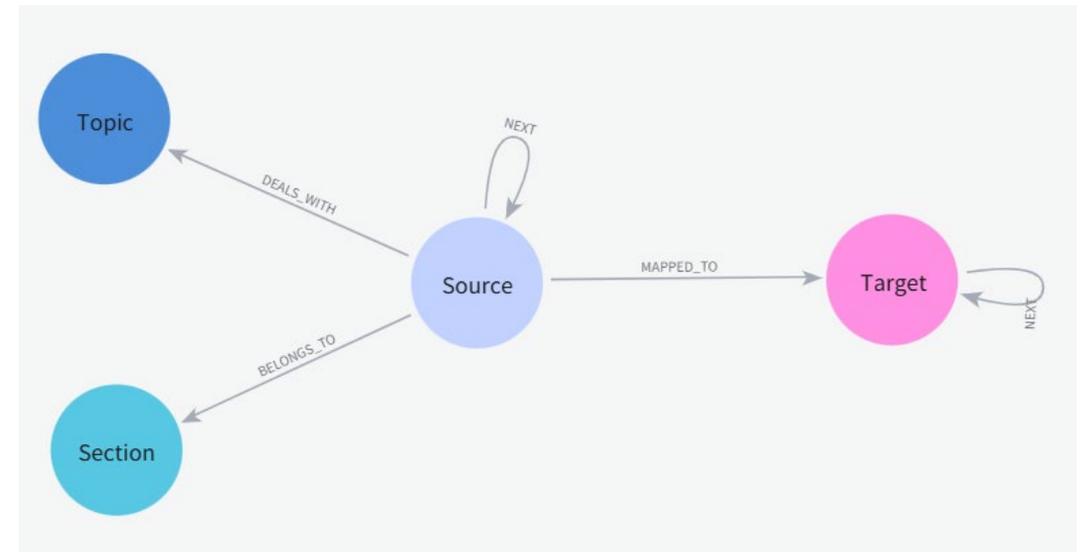
"삭제된 문장에 공통 어휘는?"

"수용된 문장의 인명은?"

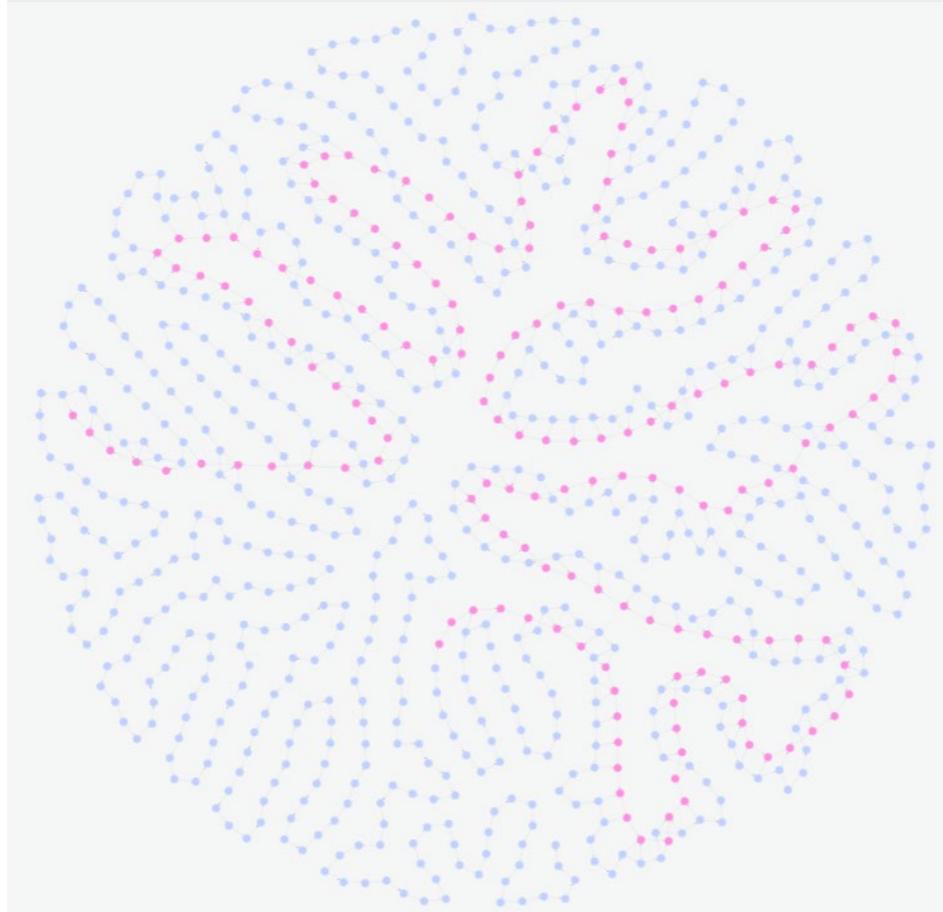
④ 시각화

과학 어휘(초록) vs 철학/권위 어휘(빨강/노랑)

[https://dh.aks.ac.kr/imk/index.php/%ED%97%88%EC%88%98\(%EA%B2%B0%EA%B3%BC-%EC%83%98%ED%94%8C\)](https://dh.aks.ac.kr/imk/index.php/%ED%97%88%EC%88%98(%EA%B2%B0%EA%B3%BC-%EC%83%98%ED%94%8C))

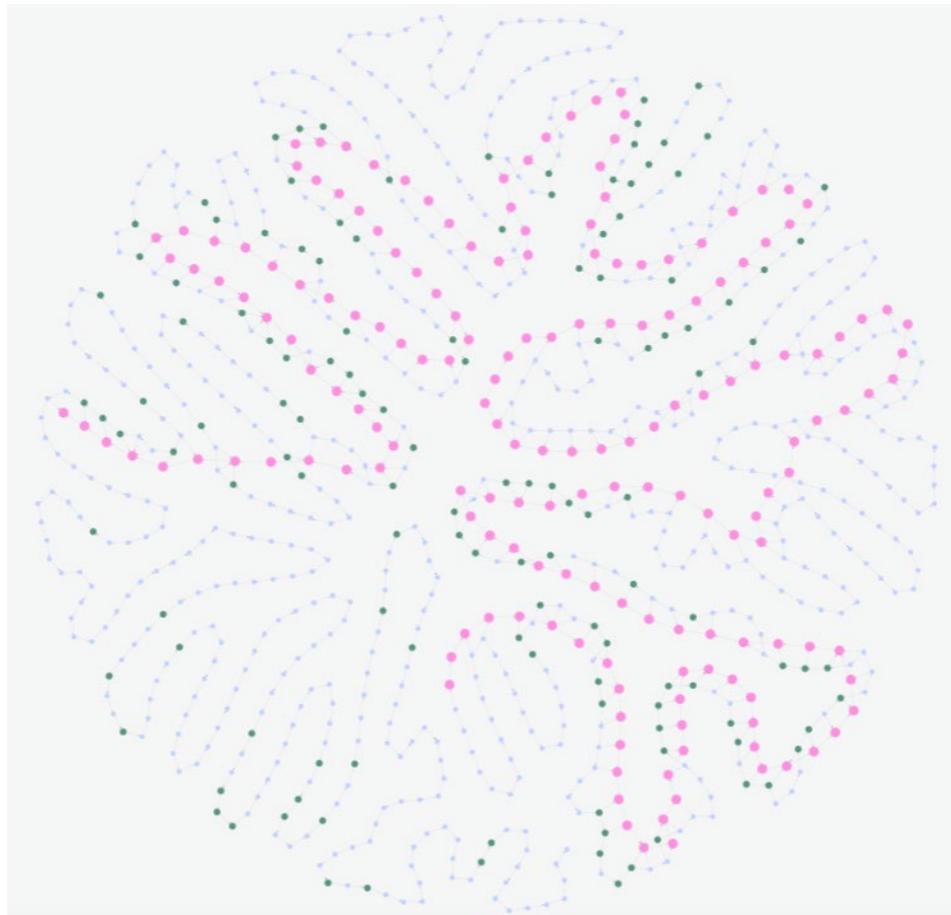


Neo4j 시각화: 전체 네트워크 구조



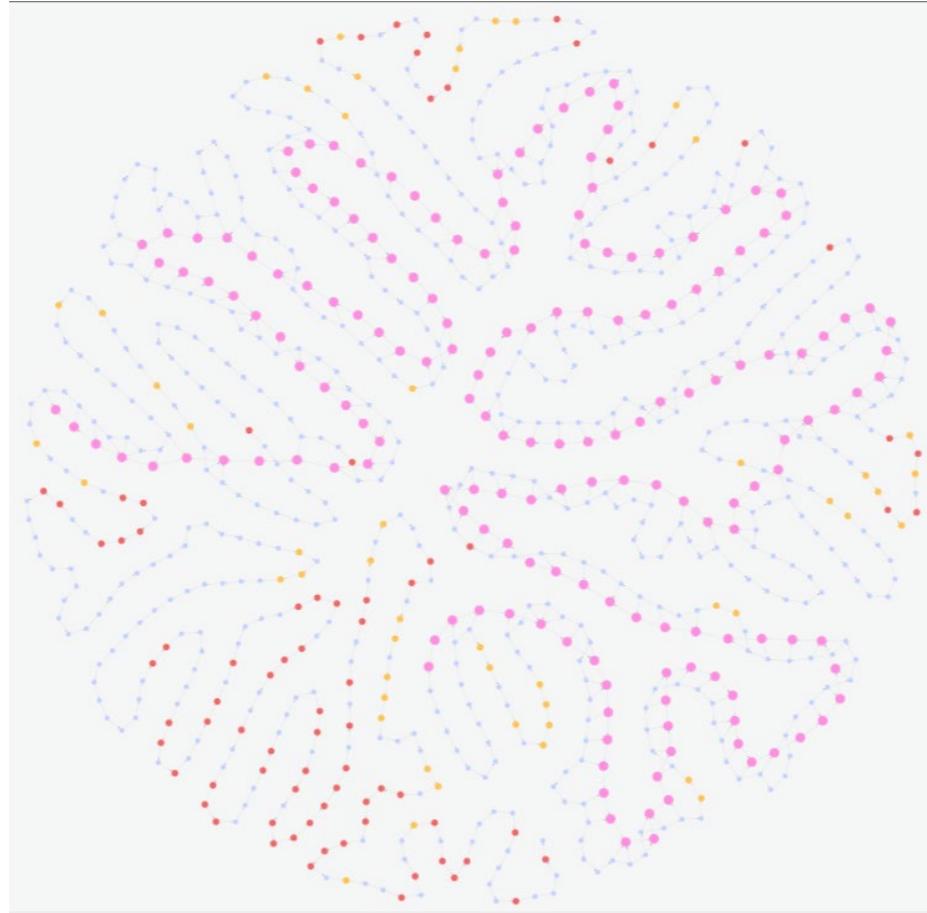
원문(689문장)과 번역문(175문장)의 관계망 — 노드: 문장/어휘, 엣지: 번역/발췌/삭제 관계

수용된 어휘: 과학 관련 (초록색 노드)



선택적으로 수용된 과학 어휘 — 생존, 생노병사, 실용적 과학 지식

배제된 어휘: 철학/권위 관련 (빨강·노랑색 노드)



삭제된 철학/권위 어휘 — 괴테, 다윈, 칸트 등 서양 철학자·과학자 인명

핵심 발견: 선택적 재구성

삭제된 것

- 서구 철학적 논의
- 서양 철학자·과학자 인명
 - 괴테, 다윈, 칸트 등
- 원문의 약 **2/3가 삭제됨**

수용된 것

- 실용적 과학 지식
- 생존·생노병사 관련 내용
- 원문의 약 **1/3만 선택** 번역

메치니코프 분석의 의의

- 단순 번역이 아닌 '**지식의 선택적 재구성**' 과정 실증
- 식민지 지식인의 **주체적 지식 편집** 양상을 데이터로 규명
- 텍스트 정독만으로는 파악하기 어려운 **구조적 패턴** 발견

【시사점】

689문장과 175문장을 일일이 대조하는 것은 매우 어려움

→ **그래프 DB**를 활용하면 구조적 패턴을 빠르게 파악

"디지털 한국사 연구는 선택이 아니라 필수이다."

왜 필수인가?

- ① **자료 환경이 이미 디지털화**되었다
연구에 사용할 자료의 대부분이 디지털 형태로 존재
- ② **연구 질문이 디지털 방법론을 요구**한다
거시적·장기적 시야를 요구하는 현재의 연구 경향
개별 정독만으로는 한계
- ③ 새로운 세대의 **자연스러운 도구**가 될 것이다
지금 시작하면 된다

어떻게 시작할 것인가?

① 두려워하지 말 것

코딩을 못해도, 수학을 못해도 괜찮다
기본 개념 이해 → 필요에 따라 조금씩 학습

② 역사학적 질문을 놓지 말 것

기술을 위한 기술이 아니다
디지털 도구는 질문에 답하기 위한 수단

③ 전통적 방법론과 상호 보완할 것

디지털이 전통을 대체하지 않는다
둘을 함께 활용 → 연구의 깊이와 넓이 확장

마무리

여러분은 지금 **역사학의 새로운 전환기**에 서 있습니다.

이전 세대가 경험하지 못한 **도구와 방법**이
여러분 앞에 놓여 있습니다.

이것을 두려움의 대상으로 볼 것인가,
기회로 볼 것인가는 여러분의 선택입니다.

적극적으로 배우고 익혀서,
한국사 연구의 **새로운 지평**을 열어가시기 바랍니다.



감사합니다