

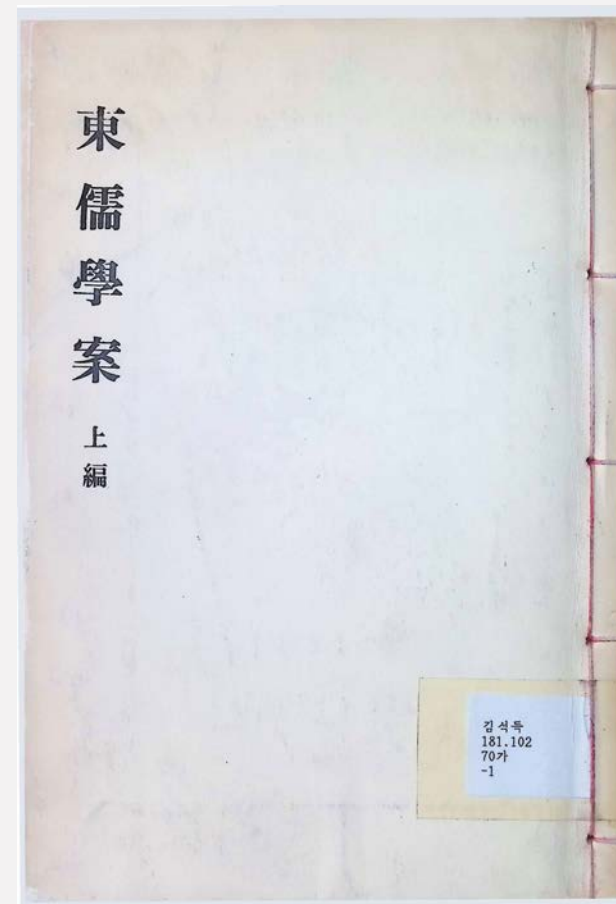
착수 발표문

『동유학안』의 디지털 표상 연구: 중층 귀속의 표상과 TEI-XML-RDF 1.2 연계

박사학위논문 중간발표

한국학중앙연구원 인문정보학과, 이재열

2026년 8월 28일 (금)



발표 구성

1부 문제 제기

연구 대상의 성격 규정과 연구 질문

2부 자료

『동유학안』의 문헌학적 고찰: 장르, 판본, 사례 분석

3부 방법론

선행 표상 모델 검토와 잠정 설계

4부 향후 계획

자료 조사, 파일럿 인코딩, 검증

PART 01

1부 문제 제기

1부 · 문제 제기 · 연구 대상

연구 대상: 하겸진 『동유학안』

편찬 개요

하겸진(河謙鎭, 1870-1946)의 조선 유학사 저작(1938-1943 편찬)

학안(學案): 학자의 전기 + 해당 학자의 저술에서 선택한 발췌문(Chu 2007, 244쪽)

체제

학파 단위 편(篇) 구성. 각 편 서두에 학파의 연원을 규정하는 총론 배치

인물 항목의 생애·학문 서술. 일부 인물에만 저술 발췌 부기

핵심 성격

사실의 저장소가 아니라 선정·분류·배열·발췌·평가로 구성된 지식 체계(최일범 2008, 11-14쪽; Wilson 1994, 4-6쪽)

연구 질문

질문 1 (핵심)

중층적 귀속 관계를 확정 사실로 기록하지 않으면서 출처 환원과 반증이 가능하도록 표상하는 방법은 무엇인가
TEI-XML 디지털 판본과 RDF 1.2 지식그래프의 구축·연계 원칙 수립

질문 2 (연장)

그렇게 구축된 데이터는 이 문헌의 학파 분류와 학술사 서술에 관하여 기존의 독해가 도달하기 어려웠던 어떠한 관찰을 가능하게 하는가

이번 발표의 위치

착수 단계 보고: 방법론 검토와 설계 방향 수립

1부 · 문제 제기 > 세 문제

기존 인물 데이터 구축의 세 문제

예시 구절: "先生始受學於家庭" (김종직 항목)

家庭이 누구인지 원문은 미명시. 인코더의 비정 필요

ex:김종직 ex:hasTeacher ex:김숙자.

문제 1: 사실화

하겸진의 서술과 인코더의 비정이 데이터에서 동일 지위로 고착

문제 2: 탈맥락화

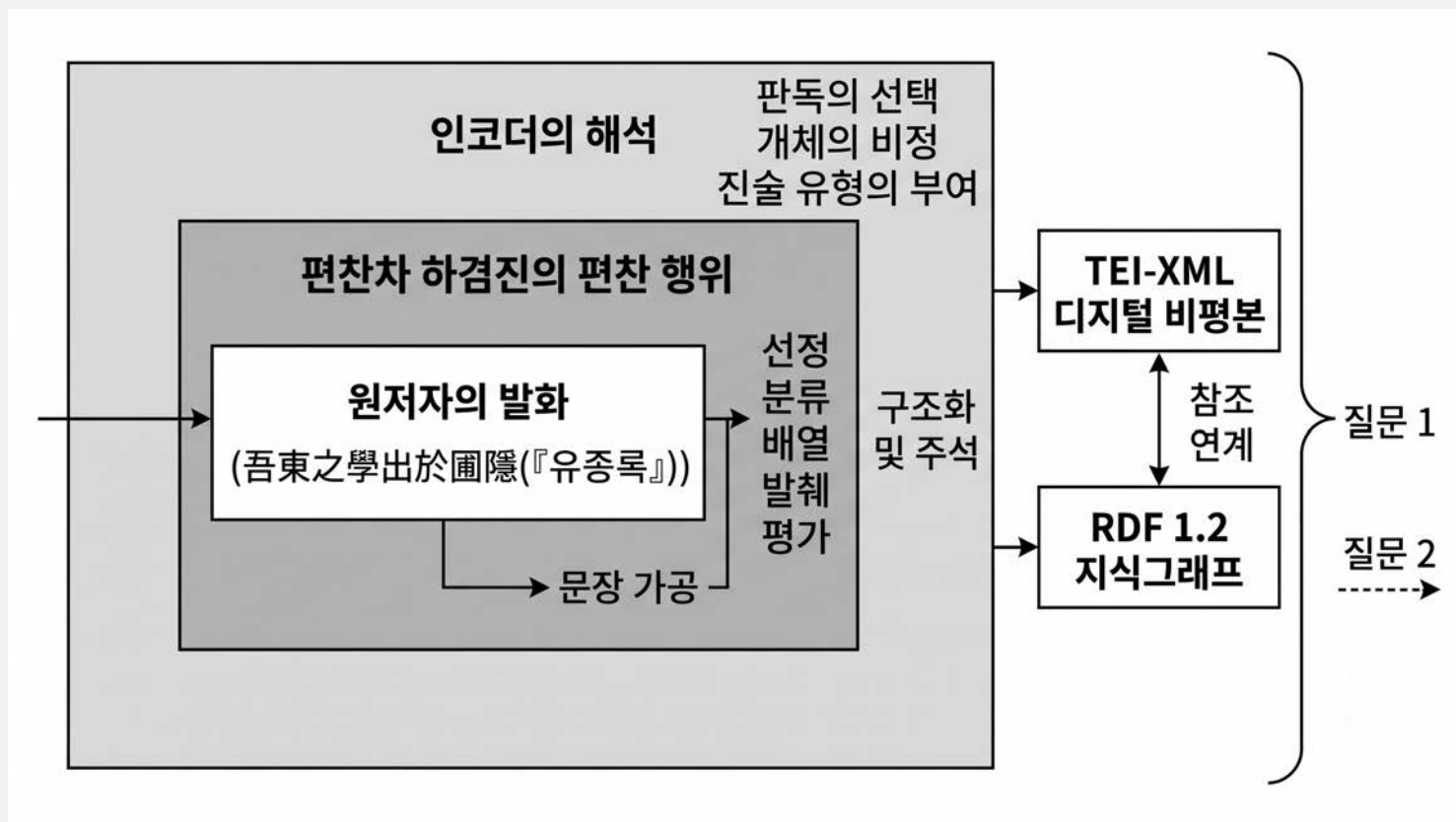
원문의 특정 구절로 되돌아갈 경로의 소실. 반증 불가능한 기록

문제 3: 중층 귀속

원저작의 발화, 편찬자의 서술과 배치, 인코더의 판단이 중첩.

단일 주체 구조는 반드시 한 층위를 삭제하게 됨.

1부 · 문제 제기 > 도해

[도해 1] 세 귀속 층위와 두 질문

1부 · 문제 제기 · 두 기준

두 기준과 연구의 성격

설계 기준

출처 환원성: RDF의 모든 진술은 판본의 구체적 위치로 환원 가능할 것

해석 책임성: 모든 판단에 판단 주체(원저자/편찬자/인코더)를 명기할 것

연구의 성격

자료 구축 연구가 아니라 방법의 성립을 입증하는 연구. 현실적 제약으로 인해 인코딩은 일부 선정 구간에 한정

산출물

TEI-XML 비판정보 + RDF 1.2 지식그래프

판본과 그래프의 통합 배제. 각자의 표준 생태계에서 구축 후 상호 참조로 연계

PART 02

2부 자료: 『동유학안』의 문헌학적 고찰

2부 · 자료 > 장르의 계보

중국 학안의 계보

저작	학안 형식 확립에의 기여	본 연구의 대응 표상 대상
『명유학안』(황종희, 1680년경)	학파별 분류의 기준 설정	학파 배속과 편차 구조
『송원학안』(황종희·전조망 외, 1838)	관계 명칭 체계를 통한 계보 표현	사승·전승 관계의 유형 체계
『국조학안소지』(당감, 1845)	명칭·배열·분량을 통한 평가의 노출	총론·평어와 배열·분량의 평가 기능

2부 · 자료 > 조선의 계보 문헌

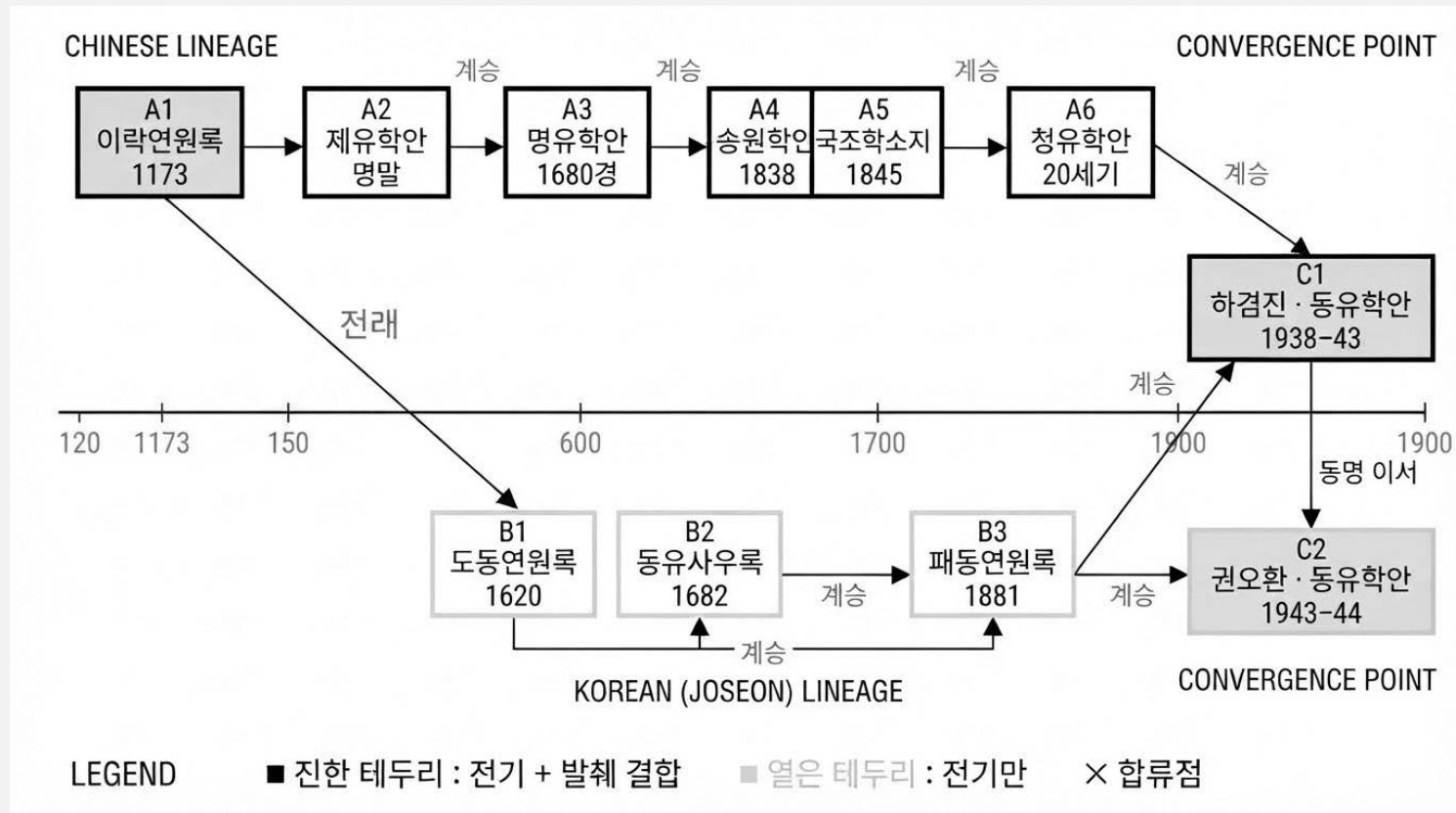
조선의 계보 문헌과 『동유학안』

구분	『도동연원록』	『동유사우록』	『패동연원록』	하경진 『동유학안』
편찬	오희길, 1620년	박세채·이세환, 1682년 완성	송병선, 1881년	하경진, 1938-1943년
기점 인물	기자	설총·최치원	정몽주	설총
수록 인원	99인	799인	69인	152인
조직 원리	도통 계보의 권별 배열	주요 인물 아래 문인·사숙·종유의 묶음	연원에 따른 인물 배열	학파별 학안과 편차
수록 자료	행장·유고	행장·묘갈명 등 행적 기록	행장·신도비명 등 타인의 기록	총론·인물 서술·본인 저술의 발췌
학설 제시	미비	미비	미비	체재의 핵심 요소

동명에서 유의: 권오환본 『동유학안』(1944)의 존재

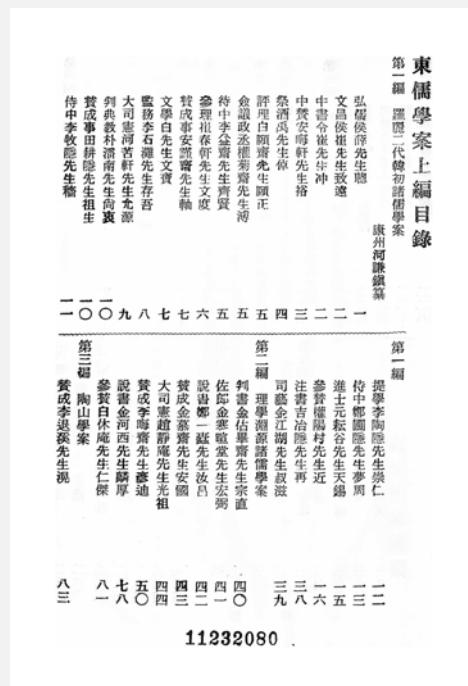
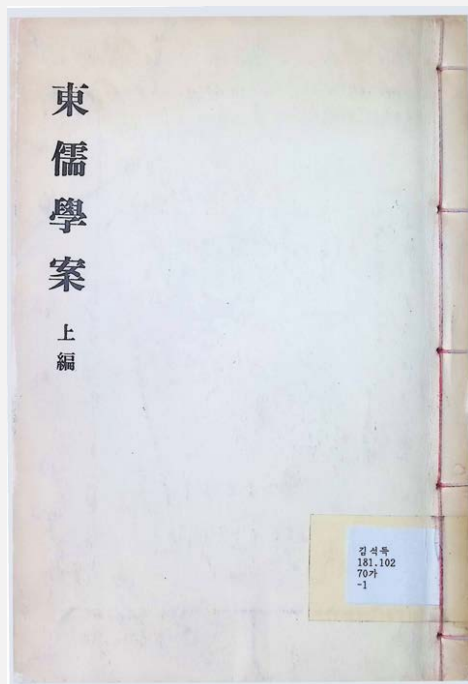
2부 · 자료 > 도해

[도해 2] 학안류와 연원록류의 편찬 연표



2부 · 자료, 판본

[도해3] 1970년 간행 『동유학안』



현재 확인되는 가장 이른 간행본. 신연활자본, 고서를 모방한 선장, 비매품

저작권발행 하영기, 편집 이일해. 문인 집단의 사업이며 발행처는 부산 해동불교역
경원 내 일봉정사

대학 도서관을 통해 3편 전량 확보, 디지털 이미지로 제작

하검진 본인의 수고본은 전하지 않음. 1970년 이전의 간행 가능성은 유보

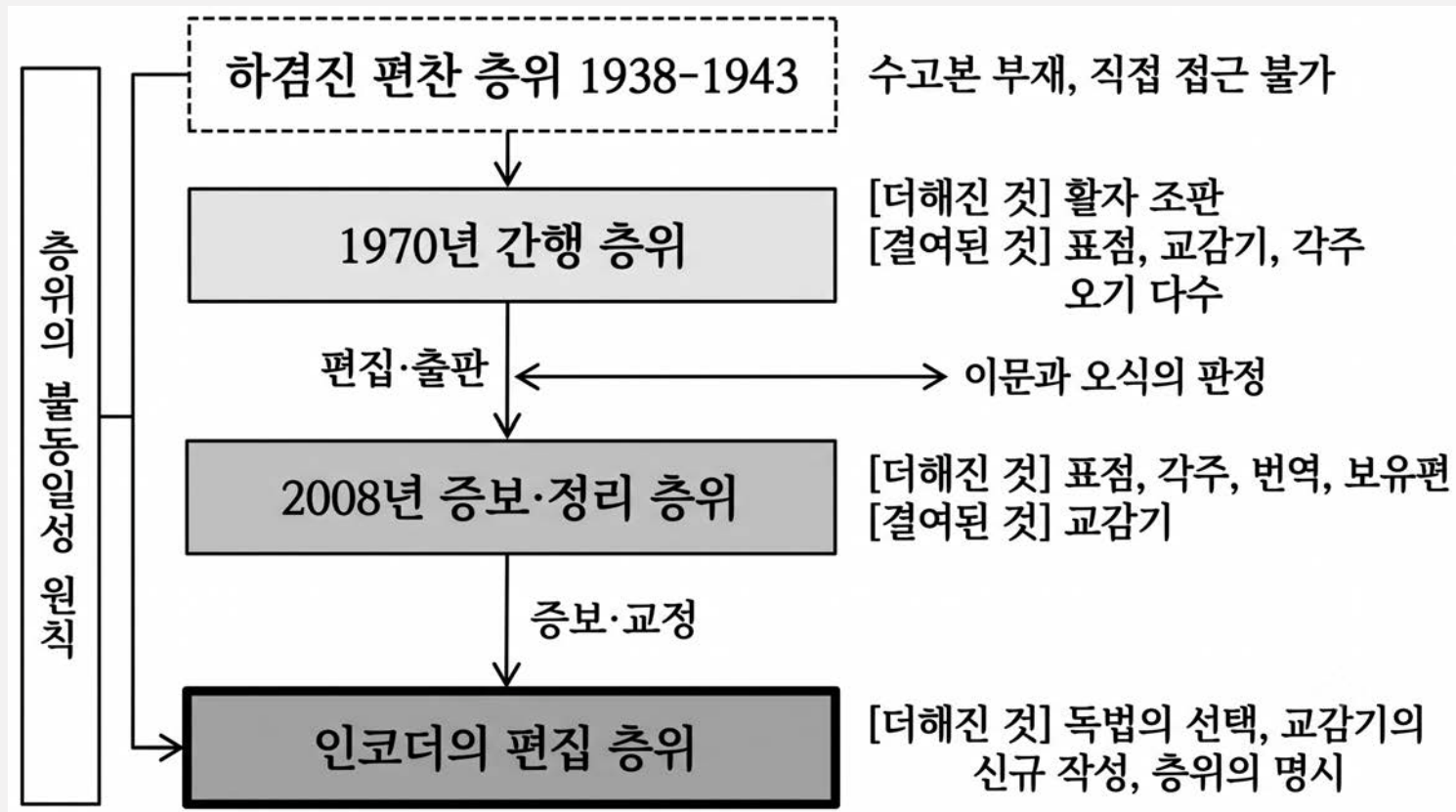
2부 · 자료 · 저본의 확정

[표] 두 판본의 대비와 저본의 확정

항목	1970년 간행본	2008년 『증보 통유학안』
체제와 수록 인원	상중하 3편 1질, 152인	6권, 23편과 보유편, 167인
표점	없음	있음
원문 교감기	없음	없음
각주	없음	교감주, 어석주, 출전 확인주, 역자 보충
번역문	없음	있음
특징적 결락	오기 다수	본문 확정 근거의 미기록
본 연구에서의 지위	대교(對校) 자료	저본(底本)
확보 범위	3편 전체(디지털 이미지 제작)	6권 가운데 1권

2부 · 자료 > 도해

[도해 4] 전승 경로와 편집 총위



2부 · 자료 · 사례 분석

[원문] 총론의 열한 진술 단위 (1): 진술 1-5

제2편 이학연원제유학안 총론 (하검진 2008, 원문편 476쪽)

1 儒宗錄有曰。吾東之學出於圃隱。圃隱傳之治隱。治隱再傳而至畢齋。自畢齋而後眞儒輩出蔚然。爲道學嫡傳。

→ 『유종록』에 이르기를, "우리 동방의 학문은 포은에게서 나왔고, 포은이 야은에게 전하였으며, 야은이 다시 전하여 필재에 이르렀다. 필재 이후로 참된 선비가 무리로 나와 성대하게 도학의 적전을 이루었다"고 하였다.

2 圃隱之爲理學祖者。盖以此也。

→ 포은이 이학의 조(祖)가 되는 것은 대개 이 때문이다.

3 寒暄一蠹學於畢齋。慕齋靜庵學於寒暄。

→ 한훤당과 일두는 필재에게 배웠고, 모재와 정암은 한훤당에게 배웠다.

4 若晦齋河西則不由師承而得者。

→ 회재와 하서 같은 이는 사승을 거치지 않고 스스로 얻은 사람이다.

5 而晦齋初學于其舅孫愚齋仲墩。河西初學於慕齋。愚齋亦畢齋之徒也。

→ 그러나 회재는 처음에 그 외숙인 우재 손증돈에게 배웠고, 하서는 처음에 모재에게 배웠으며, 우재 또한 필재의 문도이다.

2부 · 자료 > 사례 분석

[원문] 총론의 열한 진술 단위 (2): 진술 6-11과 저본의 각주

6 按退溪嘗稱畢齋爲詞華之人而不稱其道學。

→ 살피건대 퇴계는 일찍이 필재를 문사의 사람이라 일컬고 그 도학은 일컬지 않았다.

7 然其答明使許國巍時亮之問則曰道學文章爲一時儒宗。

→ 그러나 명나라 사신 허국과 위시량의 물음에 답할 때에는 "도학과 문장이 한 시대의 유종이 된다"고 하였다.

8 又按寒暄請學尙書于畢齋。畢齋手抽小學授之。

→ 또 살피건대 한훤당이 필재에게 『상서』 배우기를 청하자 필재가 손수 『소학』을 꺼내어 주었다.

9 曰。苟志於學。宜從此始。

→ 이르기를, "진심으로 학문에 뜻을 두었다면 마땅히 이것부터 시작해야 한다"고 하였다.

10 寒暄之以小學君子見稱於世。盖畢齋有以啓之也。

→ 한훤당이 세상에서 소학군자로 일컬어진 것은 대개 필재가 열어 준 바가 있었기 때문이다.

11 後之人以退溪一時之言而輕議畢齋豈知言者哉。

→ 뒷사람들이 퇴계의 한때의 말을 가지고 필재를 가벼이 논하니, 어찌 말을 아는 자이겠는가.

저본의 각주 두 개 (2008년 정리 총위)

각주 1: 진술 1의 『유종록』을 청대 인물 위상추의 저작으로 비정한 출전 확인주

각주 2: 진술 5의 손중돈에 대한 생몰년과 이력의 보충주

2부 · 자료 > 진술 단위의 분석

[표] 총론 진술 단위의 분석 (1): 진술 1-6

번호	발화 주체	진술 유형	출처 표시	표상의 쟁점
1	『유종록』의 원저자	도통 계보의 진술	서명과 도입어를 갖춘 명시적 인용	인용의 종료 지점이 원문에 표시되지 않음
2	하겸진	앞의 인용에 근거한 판정	없음	인용문의 일부인지 편찬자의 판정인지가 표점만으로 결정되지 않음
3	하겸진	사승 관계의 진술	없음	무표지 진술이 데이터에서 확정 사실로 고착되는 전형적 사례
4	하겸진	사승의 명시적 부정	없음	명시적 부정 진술을 어느 유형으로 기록할 것인지
5	하겸진	초학 관계의 진술	없음	4와의 긴장을 모순으로 볼 것인지 사승 개념의 총위 차이로 볼 것인지
6	하겸진이 전하는 퇴계의 평가	안어 안의 요약 인용	인명만 있고 출전 없음	요약 인용의 원문 범위를 특정할 수 없음

2부 · 자료 > 진술 단위의 분석

[표] 총론 진술 단위의 분석 (2): 진술 7-11과 각주

번호	발화 주체	진술 유형	출처 표시	표상의 쟁점
7	하검진이 전하는 퇴계의 발화	안어 안의 직접 인용	인명과 발화 상황은 있으나 서명 없음	원출전의 비정이 인코더의 작업으로 남음
8	하검진	일화의 서술	서명만 언급되고 인용문 없음	서명 언급을 저작에 대한 지시로 볼 것인지의 판단
9	김종직	직접 발화의 인용	화자 표시만 있고 출처 없음	출처 미상 인용의 확실성 표시
10	하검진	인과적 판정	없음	평가적 판정이 서술문의 형식을 취함
11	하검진	후대 독법에 대한 논박	없음	수사 의문으로 표현된 평가의 기록 방식
각주 1	2008년 정리 총위	출전 확인주	해당 없음	후대의 비정을 승계할 것인지 별도 총위로 기록할 것인지
각주 2	2008년 정리 총위	인물 보충주	해당 없음	저본에 없는 정보의 총위 표시

(한 단락 안에 발화의 층이 넷, 출처 표시의 정밀도는 연속적으로 교체. 김서윤 2024a의 세 등급 유형화가 기술의 좌표로 유효)

2부 · 자료 · 인용 경계

인용 경계의 두 판독

인용 종료 지점이 원문에 미표시: 두 판독은 배타적이나, 인코딩은 판정을 미룰 수 없음

```
<!-- 판독 가: 진술 2를 인용문의 일부로 보는 경우 -->
<quote source="#yujongnok">吾東之學出於圃隱。.....爲道學嫡傳。
圃隱之爲理學祖者。盖以此也。</quote>

<!-- 판독 나: 진술 2를 편찬자의 판정으로 보는 경우 -->
<quote source="#yujongnok">吾東之學出於圃隱。.....爲道學嫡傳。
</quote>
<seg type="judgement" resp="#hagyeomjin">圃隱之爲理學祖者。盖以
此也。</seg>
```

판정을 미룰 수 없음 = 판정을 인코더의 것으로 명기해야 할 이유

자료 성격에 따른 모델링 요건

1. 발화 주체 및 편찬 층위의 구분
2. 발췌된 저작의 독립 개체화
3. 판단과 사실의 구분
4. 확정과 추정의 구분
5. 텍스트 위치 참조의 다중성 수용 (원저작, 편찬물)

PART 03

3부 방법론: 선행 연구와 표상 모델의 검토

3부 · 방법론 > 요건

[표] 모델링 요건

요건	진술 요지	판정 방법
1. 발화 주체 및 편찬 층위의 구분	진술마다 누구의 발화인지를 되물을 수 있을 것	임의의 진술에 대한 발화 주체 조회가 가능한가
2. 발췌된 저작의 독립 개체화	원저작과 발췌 행위를 나누어 기술할 것	어디에서 왔는가와 누가 옮겼는가에 서로 다른 답이 조회되는가
3. 판단과 사실의 구분	평가와 분류를 사실 진술과 같은 지위로 두지 않을 것	학파 귀속 조회에 규정의 주체가 함께 조회되는가
4. 확정과 추정의 구분	불확실성을 구분하여 기록할 것	확정 진술만을 걸러 내는 조회가 성립하는가
5. 텍스트 위치 참조의 다중성 수용	하나의 단편이 두 위계(원저작, 편찬물)에 속함을 지시 가능할 것	발췌문에서 두 위계의 좌표를 모두 얻을 수 있는가

3부 · 방법론 > 팩토이드

팩토이드: 사실에서 진술로

팩토이드(factoid)의 정의

사실(fact)이 아니라 사료가 주장하는 바의 기록(factoid)

자료 S의 위치 L에서 인물 P에 대해 F를 진술한다는 주장(Bradley and Short 2005, 7-8쪽)

모순 진술의 병존 허용 = 핵심 질문에 대한 최초의 구조적 해답

사료 주도적(source-driven) 성격

역사학자의 독해가 아니라 사료 자체를 우선(Pasin and Bradley 2015, 89쪽).

온톨로지 재정식화로 **해석 주체**가 별개 개체로 데이터에 진입

본 연구와의 관계

단일 편찬물 내부의 중층 귀속 → **문헌의 어느 지점이 누구의 목소리로 무엇을 말하는가의 기록** 요구

3부 · 방법론 > 구조 대비

사실화 트리플과 팩토이드 구조의 대비

김종직 사승관계의 두 기록 방식: "先生始受學於家庭"의 家庭을 김숙자로 비정하는 경우

```
# 사실화 트리플: 두 층위(출처, 해석 행위)가 누락됨
ex:김종직 ex:hasTeacher ex:김숙자 .
```

```
# 인코더, 출처문서, 해석 결과, 그리고 해석 행위를 분리
ex:Encoder_E a ex:Encoder .
ex:Dongyuhagan a ex:Document .
ex:TeachingEvent_001 a ex:TeachingEvent ;
    ex:teacher ex:김숙자 ;
    ex:student ex:김종직 .
ex:Interpretation_001 a ex:DocumentInterpretationAct ;
    ex:author ex:Encoder_E ;
    ex:source ex:Dongyuhagan ;
    ex:interprets ex:TeachingEvent_001 ;
    ex:evidencePointer ex:DYH_tei_anchor_001 .
```

위 트리플에서 누락된 두 층위

이 트리플이 『동유학안』이라는 문헌에서 비롯되었다는 차원. 인코더가 해당 구절을 그렇게 읽었다는 차원

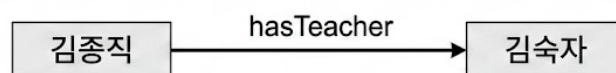
아래 구조가 확보하는 것

누가 어떤 근거로 구조화했는지의 명시. 동일 구절에 대한 다른 해석의 대안 해석으로서의 수용(오류가 아닌 병존)

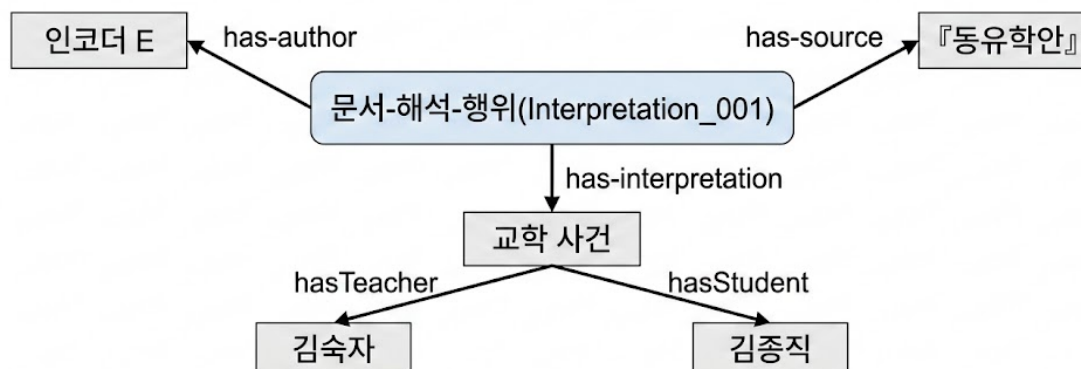
3부 · 방법론 > 도해

[도해 5] 팩토이드 모델 개념도

단순 트리플 방식



팩토이드 모델



3부 · 방법론 > FPO

FPO와 이중 매개 구조의 공백

팩토이드 온톨로지(FPO)에 의한 같은 사례의 기록

```
ex:factoid_001 a fpo:RelationshipFactoid ;
    fpo:sourcedFrom ex:citation_001 ; fpo:hasReference ex:ref_001 , ex:ref_002 .
ex:citation_001 a fpo:SourceCitation ;
    fpo:fromSource ex:source_A ; fpo:hasPlaceInSource "제2편, 김종직 항목" .
ex:source_A a fpo:Source ; fpo:hasSourceName "東儒學案" .
ex:ref_002 a fpo:PersonReference ;
    fpo:hasNameInSourceAs "家庭" ; fpo:referencesPerson ex:person_kimsukja .
```

이중 매개 구조의 성과

사료 속 표기(家庭)와 비정된 인물(김숙자)의 분리(PersonReference), 진술과 출처 위치의 분리(SourceCitation)

한계

:Source에 『동유학안』을 적는 순간, 그 구절이 하겸진의 저작인지 옮겨 온 인용인지를 되물을 자리 부재

분리되는 것은 사료의 층위와 인코더의 층위 둘뿐. 사료 안쪽은 미분리

3부 · 방법론 > 판본학 계보

두 계보의 공백과 단일 귀속의 손실

디지털 판본학 계보의 응답

단언적 판본(assertive edition, Vogeler 2019)

편찬자의 해석 = 판본 데이터의 구성 요소.

RDF 요소를 TEI-XML 구문 내에 포함시킬지는 미결

판본 도메인 온톨로지(Boot and Koolen 2021)

저작(work)/문헌(document)의 구별, 교차점의 개체화

공통의 공백

단일 발화 주체 + 단일 해석 주체. 사료 내에 존재하는 편찬자의 자리 부재.

FPO의 확장 사례

시간 차원의 확장 사례(小川潤·大向一輝 2021)

→ 동아시아 사례에서 FPO의 확장 가능성을 보여줌. 귀속 차원의 확장은 본 연구의 고유 과제

단일 귀속 구조에서 잃는 세 가지

편찬의 행위가 기록되지 않는다. 선택과 배치가 진술 내용에 흡수된다.

발화의 주체가 뒤바뀐다. 『유종록』의 진술이 하겸진의 주장이 되며, 이는 학술사적 오류이다.

인코더의 판단이 한 종류로 줄어든다. "家庭"을 김숙자로 읽는 인물 동일시는 기록되지만, 이 구절이 인용이라는 판단이 기록되지 않는다.

3부 · 방법론 > 어휘

단일 귀속 구조 해결을 위한 어휘 (1): 논증의 어휘

논증의 어휘 (CRMInf): 누가 무엇을 주장했는가를 기록 (Doerr 외 2026)

CIDOC CRM의 확장, 명제 자체가 아니라 **명제를 주장하거나 받아들인 행위**를 과거에 실제로 일어난 사건으로 기록

지식 생성의 경로를 **관찰**과 **믿음 채택**과 **추론 수행**으로 분류. 편찬자와 인코더의 행위를 서술 가능.

데이터가 단언하는 바는 조선의 학문이 포은에게서 나왔다는 것이 아니라, **하경진이 『유종록』을 근거로 그 명제를 받아들였다는 것**

참이라는 표시는 명제가 옳다는 뜻이 아니라 **그 행위자가 그렇게 받아들였다는 뜻**. 미상 값 표현 가능.

→ **발화의 주체가 뒤바뀌지 않고, 인코더의 판단이 대상과 확신도와 논리로 나뉘어 기록된다**

3부 · 방법론 > 어휘

단일 귀속 구조 해결을 위한 어휘 (2): 출처와 주석의 어휘

출처의 어휘 (PROV-O): 무엇이 무엇에서 나왔는가를 기록

어떤 자료가 어떤 활동을 거쳐 어떤 자료에서 파생되었는가를 기록하는 W3C 표준

→ 편찬자가 인용해 왔다는 행위 자체를 표현 가능

주석의 어휘 (Web Annotation Model): 어디에 관한 말인가를 기록한다

주석을 독립된 개체로 세우고 말하는 내용과 그 대상을 분리하는 W3C 표준

하나의 주석이 여러 대상을 동시에 가질 수 있고, 대상의 어느 부분인지는 여러 방식의 지시자로 지정

발췌된 인용문이 원저작의 위계와 편찬물의 위계에 동시에 속한다는 사태를 표현 가능

→ 위치 참조의 다중성이라는 요건에 응답

3부 · 방법론 > 대응표

[표] 요구사항 대응표

이 표는 모델의 우열이 아니라 기존 모델 위에 무엇을 추가해야 하는가를 보기 위한 것임

모델	발화 주체 구분	저작의 개체화	편찬자 판단의 표시	불확실성 표시	위치 참조의 다중성
팩토이드와 FPO	△ 사료 안쪽 미분리	×	×	△ 확장에서 제공	×
단언적 판본의 RDF 내장 TEI	△ 발화 주체 하나	×	△ 지위 구분 장치 없음	△ TEI 속성 의존	△ 층위가 한 지점에 중첩
판본 도메인 온톨로지	×	△ 발체 행위의 자리 없음	×	×	○ 교차 단편의 개체화
CIDOC CRM과 Bio CRM	×	△	△	×	×
CRMinf	△ 텍스트 위치와 미연결	×	○ 믿음·믿음값	○ 추론·논리 기록	×
PROV-O	△ 파생 이력으로 표현	△ 발체 활동 기술	△	×	×
Web Annotation Model / 나노퍼블리케이션	△ 작성·반입·주장 삼분	×	△	×	○ 다중 대상·선택자

3부 · 방법론 > 개체화

인용 행위의 개체화

두 방식의 대비: 속성 부여 vs 개체화 (본 연구는 후자 선택)

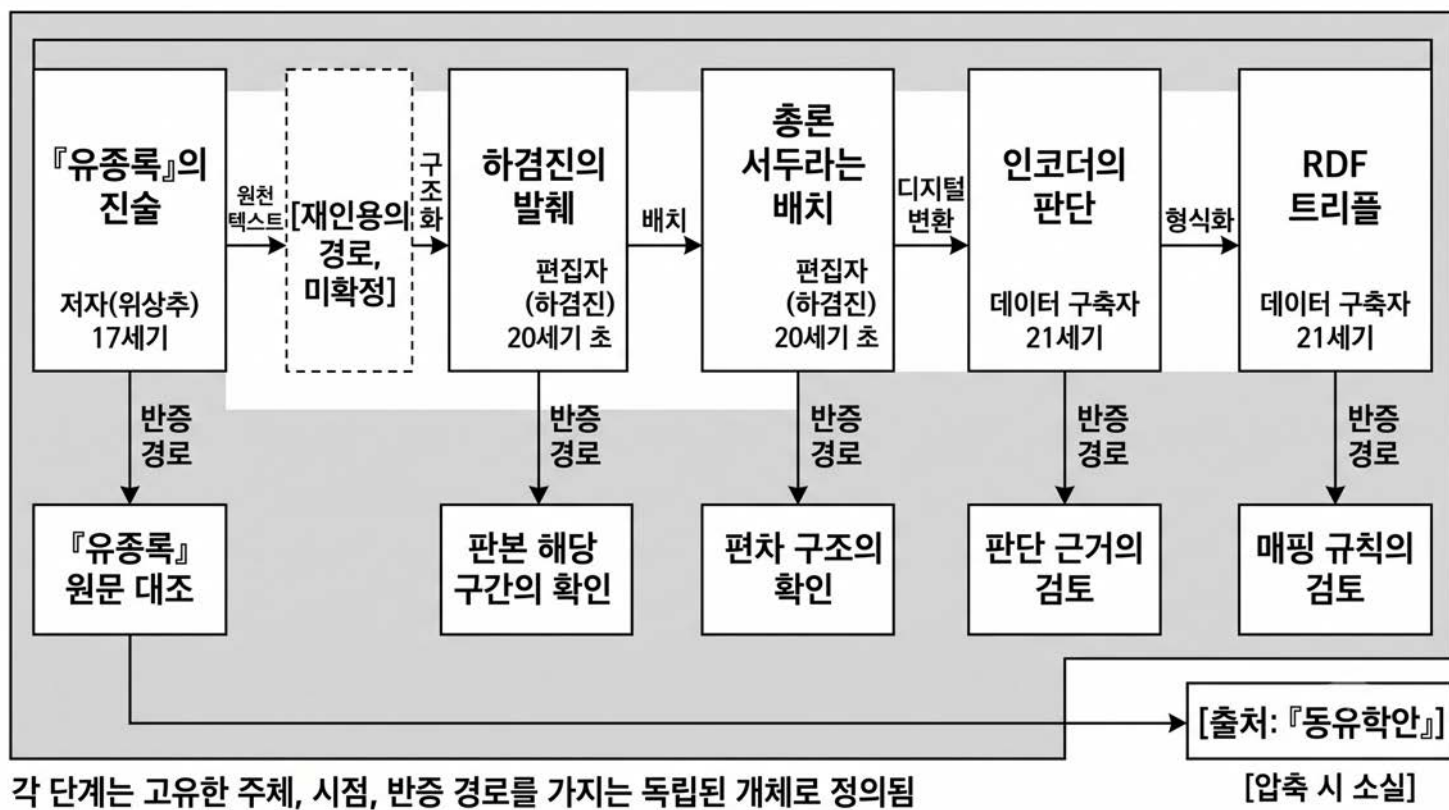
```
<!-- 전자: 인용문 요소에 분류값을 속성으로 부여 -->
<quote xml:id="q-yujongnok-01" source="#yujongnok"
      ana="#citeInfo-full #citeForm-verbatim #citePurpose-authority">
  吾東之學出於園隱。.....爲道學嫡傳。</quote>
```

```
# 후자: 인용 행위를 개체로 세우고 분류를 그 개체에 관한 주장으로 기록
ex:citing_001 a cito:Citation ;
  cito:hasCitingEntity ex:prologue_ch2 ;
  cito:hasCitedEntity ex:work_yujongnok ;
  cito:hasCitationCharacterization cito:citesAsAuthority ;
  ex:evidence ex:citation_476 .

ex:judgement_001 a crminf:I2_Belief ;
  rdf:reifies <<( ex:citing_001 cito:hasCitationCharacterization
                  cito:citesAsAuthority )>> ;
  prov:wasAttributedTo ex:person_encoder .
```

3부 · 방법론 > 도해

[도해 6] 5단계 귀속 연쇄



3부 · 방법론 > 편집 판단

판본의 편집 판단과 출처 정보의 내장 (1)

Kuczera, Andreas. "TEI Beyond XML - Digital Scholarly Editions as Provenance Knowledge Graphs." In *Graph Technologies in the Humanities - Proceedings 2020*, edited by Tara Andrews, Franziska Diehr, Thomas Efer, Andreas Kuczera, and Joris van Zundert. 2020.

전사자·편집자·주석자·이용자를 **판본 체계 바깥의 관찰자**가 아니라 **그 체계의 구성 요소**로 인정(Kuczera 2020, 102쪽)

"해석 과정은 텍스트의 첫 문자가 전사되는 순간부터 시작된다. 편집자(editor)의 각 결정은 논의의 여지가 있는 주관적 진술이며, 바로 이 지점에서 **출처 정보**가 개입한다." (Kuczera 2020, 115쪽)

귀결: 디지털 판본 전체를 출처 정보를 내장한 지식그래프로 볼 것

판본 = 참 명제의 모음집이 아니라 각기 다른 주체에게 귀속되는 진술들의 집합

판독·표점·교감·주석, 나아가 주석에 대한 주석까지 모두 진술. 진술마다 그 발화자와 시점이 추적 가능해야 함

3부 · 방법론 > 편집 판단

판본의 편집 판단과 출처 정보의 내장 (2): 계승과 이견

계승

인코더의 편집 판단을 데이터의 일부로 기록한다는 전제. 2부에서 세운 편찬 층위의 비동일시 원칙과 같은 자리

이견

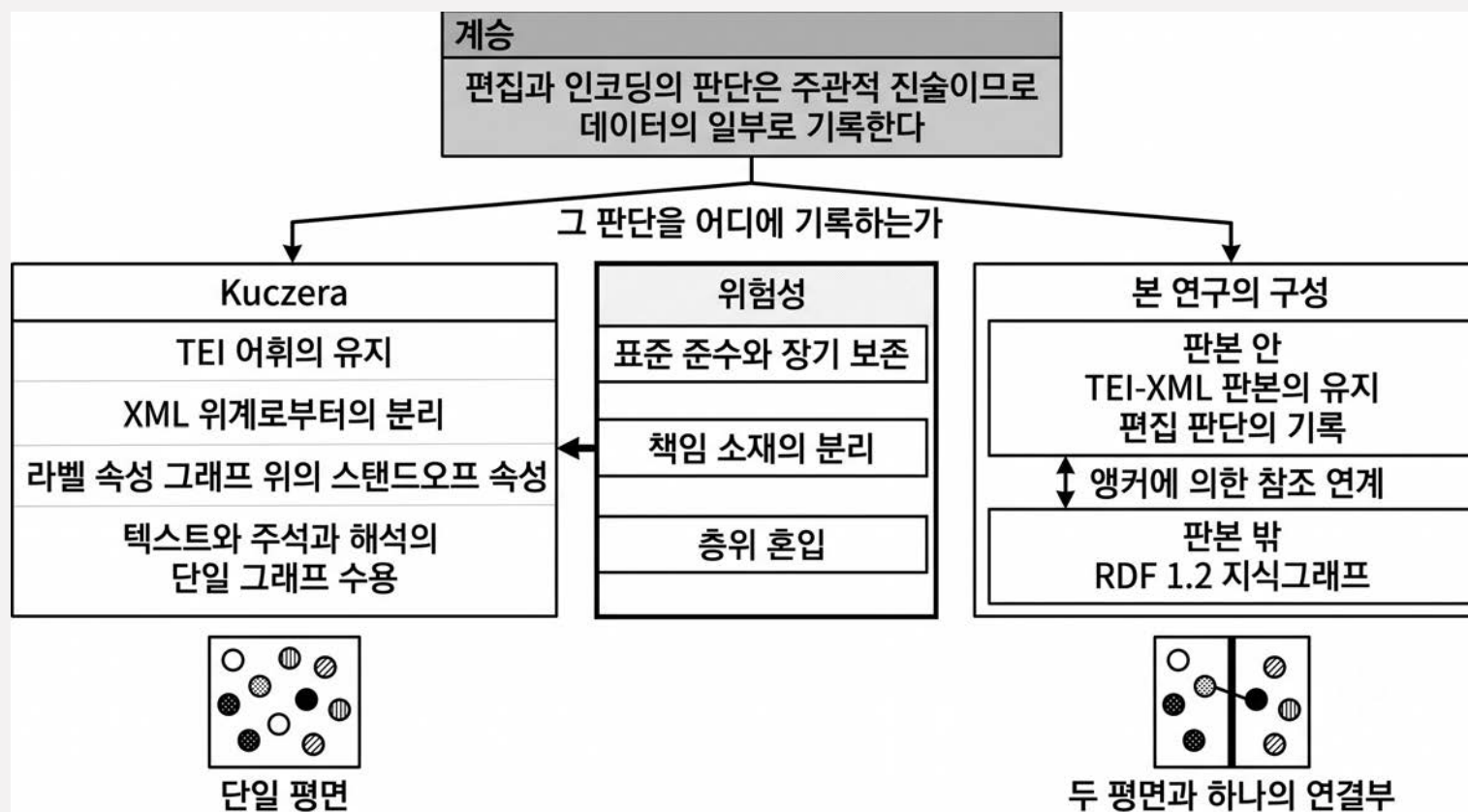
Kuczera는 TEI 어휘는 유지하되 이를 XML에서 분리하여, 라벨 속성 그래프(LPG=Neo4J) 위에서 스탠드오프 속성(SPO)으로 기술하자고 주장 → 표준 준수와 장기 보존, 편집학적 판단과 모델링 판단의 책임 소재 분리, 층위 혼입의 위험

본 연구의 방향

판본 자체는 TEI-XML로 유지. 출처 정보 내장은 판본 내의 편집 판단 기록과 판본 바깥의 지식그래프의 참조 연계로 분담

3부 · 방법론 > 도해

[도해 7] 디지털 판본과 지식그래프의 관계 설정



3부 · 방법론 > 데이터 모델

그래프 데이터 모델의 선택: RDF 1.2

라벨 속성 그래프(Labeled Property Graph, Neo4J)의 장점

진술에 대한 메타데이터 부착이 모델의 기본 기능

표현의 경제성: RDF보다 짧은 구문에 동일 정보량을 담을 수 있음

ISO GQL로 질의 언어의 공적 표준화 진전

RDF 1.2 잠정 선택의 근거

기존 RDF 1.1 표준 구체화(reification) 방식에 비해 메타데이터 부착이 용이해짐(RDF-star에 비하면 경제성이 떨어지나 표현 정밀도 증가).

기존 어휘체계와의 호환성: CIDOC CRM, Bio CRM, CRMInf, FPO 등 이용 가능

전역 식별자(IRI)

LPG 모델은 핵심 스토리지 엔진 구현과 생태계 주도권이 특정 상용 기업에 크게 의존

[주의사항] 미결정 표준으로서의 위험성. SPARQL 1.2는 작업 초안. 2026년 8월 현재 저장소 지원 불균등.

3부 · 방법론 > 표기 방법

진술 수준 메타데이터의 표기 방법

동일 명제(김종직의 도학 적전 배치)에 대한 세 표기

```
# 1) RDF 1.1 표준 구체화(standard reification)
ex:stmt_001 a rdf:Statement ;
  rdf:subject ex:kimjongjik ; rdf:predicate ex:placedIn ;
  rdf:object ex:orthodoxLineage ;
  ex:statedIn ex:yujongnok ; ex:adoptedBy ex:hagyeomjin .
```

```
# 2) RDF 1.2 구체화: 같은 명제에 복수의 구체화자(reifier)
ex:r_yujongnok  rdf:reifies <<( ex:kimjongjik ex:placedIn ex:orthodoxLineage )>> ;
  a ex:SourceStatement ;  ex:statedIn ex:yujongnok .

ex:r_hagyeomjin rdf:reifies <<( ex:kimjongjik ex:placedIn ex:orthodoxLineage )>> ;
  a crminf:I2_Belief ;    ex:adoptedBy ex:hagyeomjin ;
  ex:evidence ex:citation_476 .
```

3부 · 방법론 > 이원 구축

이원 구축과 판본 내 자기 기술

연계의 세 유형(그래프 전환 / TEI 내장 / 참조 연계) 가운데 셋째를 택하되 둘째의 산출물을 연결부로 흡수

```
<TEI>
  <teiHeader>
    <xenData>
      <rdf:RDF>
        <edd:LocatedTextFragment rdf:about="#q-yujongnok-01">
          <edd:fragmentOfWork rdf:resource="work/yujongnok"/>
          <edd:locatedInDocument rdf:resource="donguhagan/ed2008#p476"/>
        </edd:LocatedTextFragment>
      </rdf:RDF>
    </xenData>
  </teiHeader>
  <text>
    <quote xml:id="q-yujongnok-01" source="#yujongnok">吾東之學出於圃隱。.....</quote>
  </text>
</TEI>
```

분업 구조

본문 요소에는 식별자만. 판본 정보는 헤더에 집약. 해석(주장의 총위)은 판본 외부의 지식그래프에 배치

#q-yujongnok-01 = 두 산출물이 만나는 주요 접점

3부 · 방법론 > 앵커

3종의 지시자와 이중 앵커 전략 (1)

기준	구조 기반	위치 기반	내용 기반
개정 내구성	높음	낮음(한 글자 증감에 취약)	중간(교감·표점 변경에 취약)
지시 입도	판본 인정 단위에 한정	문자 단위 임의	문자 단위 임의
구현 여건	넓음	좁음(TEI 스킴 파서 결여)	넓음(주석 표준 선택자)

이중 앵커 전략: 구조 앵커(structural anchor)에 증거 앵커(evidential anchor)를 종속 구성

구조 앵커(structural anchor): 편, 학안, 인물 항목, 인용문처럼 판본 구성 단위에 부여하는 고정 주소

증거 앵커(evidential anchor): 그 단위 안에서 특정 주장의 근거가 되는 문장이나 구절의 범위를 짚는 정밀 지시

3부 · 방법론 > 앵커

3종의 지시자와 이중 앵커 전략 (2): 이유

구절 단위 지시는 근거 확인에 유리하나 글자 하나의 변경에도 어긋나고, 큰 구조 단위 지시는 오래 유지되나 어느 대목이 근거인지 알 수 없음

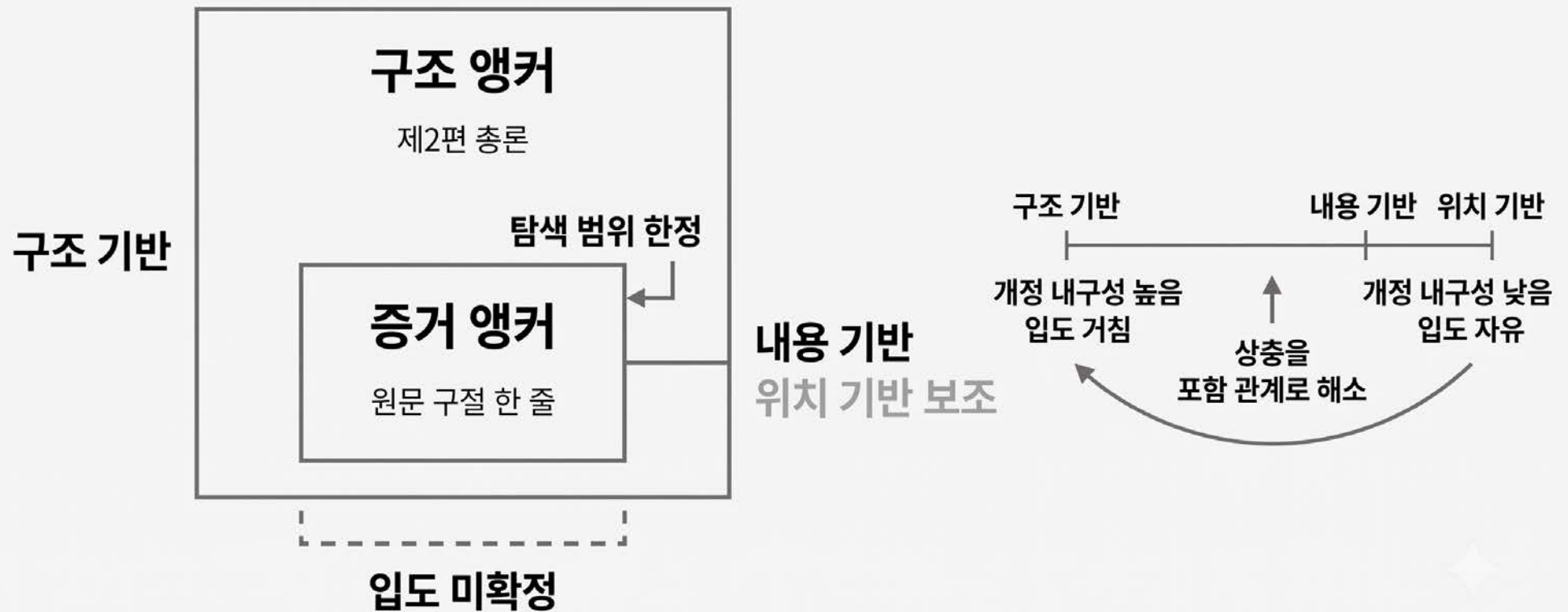
개정 내구성이 높으나 입도(granularity)가 거친 계열을 상위에, 입도는 자유로우나 취약한 계열을 하위에 배치

상위 앵커가 하위 앵커의 탐색 범위를 미리 한정하므로, 표점 수정으로 증거 앵커가 어긋나도 어디를 고쳐야 하는지 쉽게 찾을 수 있음

앵커를 어느 정도로 세분할 것인가는 실제 작업 이후 결정

3부 · 방법론 > 도해

[도해 8] 이중 앵커 구조



PART 04

4부 향후 계획

4부 · 향후 계획 > 파일럿

파일럿 인코딩과 검증

파일럿 구간의 선정 기준

학파적 대표성이 아니라 모델 검증에 용이한 사례

인코딩의 네 국면

판본 층위의 편집 → 식별 층위(개체·비정 분리) → 진술 층위(다섯 단계 구체화자) → 증거 층위(이중 앵커)

검증

시범 질의 시나리오(요구사항별 판정 질의). 앵커 내구성 시험(수정 후 지시 유지의 계량).

병행 산출물

편집 지침 초안 = 방법론의 자체 시험

생성형 AI 시험 (Pollin 외 2025)

4부 · 향후 계획 > 맺음말

맺음말

이 연구가 입증하려는 것

특정 문헌의 데이터가 만들어졌다는 사실이 아니라, 편찬된 지식 체계를 확정 사실로 환원하지 않으면서 표상하는 방법의 성립

설계의 요체

출처 환원성과 해석 책임성 → 팩토이드 골격의 귀속 차원 확장

→ 이원 구축(TEI-XML+RDF)과 이중 앵커(구조+증거)

감사합니다