

『상서고훈』(尙書古訓)과 『상서고주』(尙書古注) 연계 시맨틱 데이터 프로세싱 2 온톨로지 디자인과 파싱

변은미*
이동학**

초록 본 논문은 다산 정약용의 『상서고훈』(尙書古訓)과 석천 신작의 『상서고주』(尙書古注) 연계 시맨틱 데이터 프로세싱의 두 번째 연구로서, 기존 XML 기반의 반정형 데이터를 온톨로지 기반의 트리플 데이터(Triple data)로 전환하는 과정을 다룬다. 이를 위해 경학 자료의 구성적 특징과 주요 문맥 정보 및 유관 해석 정보들을 바탕으로 의미적 연결 관계에 중점을 둔 온톨로지를 설계하고, XML 데이터 파싱 과정 및 지식 그래프 구현 사례를 제시함으로써 경학 자료에 내재한 해석 논리와 주석 간 상호작용을 입체적으로 분석하기 위한 방법론적 기반을 마련하고자 하였다.

XML 스키마와 온톨로지 디자인이 전체적인 구조적 틀에서 설계 논리를 공유한다는 점을 바탕으로, 자료의 형식 정보, 문맥 정보, 해석 정보를 포괄하는 개념적 구조를 정의하였다. 이어서 정보 간 세부 연계 방식의 차이를 반영하여 총 15건의 클래스를 설정하고, 클래스 간 관계 및 그 연계적 의미를 고찰하였다. 특히 주석 간 관계 유형을 기본형, 내포형, 종속형으로 세분화함으로써 인용 간 관계와 주석 내부의 논리 구조를 체계적으로 반영하고자 하였으며, 관계 속성을 활용한 문맥 정보의 연결을 통해 자료 내 의미적 연결성과 해석의 계보성을 구체화하였다.

파이썬(Python)과 LLM을 활용하여 XML 데이터를 파싱하고 각 클래스의 노드(Node)와 엣지(Edge)를 구성하는 데이터를 추출함으로써 그래프 데이터베이스 구축의 기반을 마련하였다. 또한 지식 그래프를 일부 구현하여 경학 자료의 해석 구조를 시각화함으로써, 경학 자료를 대상으로 한 의미 기반 정보의 연결 방안과 시맨틱 데이터의 활용 가능성, 자료 간 비교 연구로의 확장 가능성 등을 실험적으로 검토하였다.

주제어 경학, 상서, 상서고훈, 상서고주, 정약용, 신작, 디지털 인문학, 시맨틱 데이터, 온톨로지, 지식 그래프, XML 파싱, 트리플 데이터

* 고려대학교 국어국문학과 박사수료, 제1저자

** 고려대학교 국어국문학과 박사수료, 제1저자

1. 머리말: XML 데이터 가공과 트리플 데이터 편찬

본 연구는 다산 정약용의 『상서고훈』(尙書古訓)과 석천 신작의 『상서고주』(尙書古注)를 대상으로 시맨틱 데이터 프로세싱을 수행함으로써, 경학 자료에 내재한 다층적 지식 관계망을 종합적으로 재현하고 해석 논리 및 주석 간 상호작용을 분석할 수 있는 데이터 모델을 제시하는 데 목적이 있다. 연구의 첫 단계에서 두 자료의 형식적-내용적 특성을 반영한 XML 문서를 편찬하였고, 이 내용을 담은 첫 번째 논문을 발표한 바 있다.¹

본 논문은 시맨틱 데이터 프로세싱의 두 번째 단계로서, 『상서고훈』과 『상서고주』의 XML 데이터를 온톨로지(Ontology)와 연계하여 그 의미와 맥락을 부여하는 과정을 다룬다. XML 설계와 온톨로지 디자인은 절차적으로는 순차대로 수행되었지만, 실제 작업에서는 두 과정이 유기적으로 맞물려 진행되었다. XML 스키마는 경학 자료의 분절 단위, 자료 고유의 영역 구분, 인용과 견해의 제시 방식 등 원자료의 형식적 특성을 충실히 반영하는 것을 기본 원칙으로 하되, 온톨로지 설계를 염두에 두고 의미적 연결 가능성이 있는 요소를 중심으로 체계화하였다. 동시에 온톨로지 디자인은 XML 문서의 핵심 구성 축과 구조적으로 대응하면서, 경학 자료를 구성하는 정보들이 맺고 있는 관계를 트리플(Triple, S-P-O) 형식의 데이터로 기술하기 위해 클래스(Class), 속성(Attribute), 관계(Relation)를 정의하는 방식으로 설계되었다. 이처럼 XML 스키마는 온톨로지 설계의 기준으로 작용하였고, 온톨로지의 분류 체계는 XML 구조의 세부 정비에 영향을 미침으로써 두 작업은 상호보완적으로 수행되었다.

XML 데이터를 온톨로지 기반의 트리플 데이터로 가공하는 작업은 데이터에 내포된 인문학적 의미를 더욱 구체화하고 명시화하는 과정이다. 연구의 첫 단계에서 편찬한 XML 문서는 원자료의 형식과 내용을 세밀하게

1 변은미·이동학·유인태(2024), pp. 287-338.

반영하는 동시에 연구자가 본격적으로 연구에 활용할 수 있도록 여러 가지를 고려하여 설계되었다. 다만 XML 문서만으로는 요소 간의 의미적 관계나 개별 정보의 다양한 맥락을 충분히 드러내기가 어렵다. 특히 경학 자료와 같이 주석 내 인용 관계망이 복잡하고 주석자의 사유 구조가 중첩적으로 전개되는 경우에는 마크업만으로 지식의 구조를 온전히 기술하는 데 한계가 있다. 이에 본 연구에서는 XML 데이터를 온톨로지 기반의 트리플 데이터로 전환함으로써, 텍스트 내 개별 요소들에 의미 기반의 식별성과 관계성을 부여하고자 하였다.

이와 관련하여 2장에서는 온톨로지 디자인의 기본적 구조를 소개하고 XML 스키마와의 연계 방식을 다룬다. 첫 번째 논문에서 제시한 ① 원문요소, ② 주석요소, ③ 문맥요소, ④ 해석요소의 네 가지 항목을 중심축으로 계승하되, 그중 의미적 연결 가능성을 지닌 항목들을 선별하여 범주화하고 이를 온톨로지의 구성 정보로 조직하는 방안을 중점적으로 논의하였다. 그 과정에서 XML 데이터의 보완 지점을 확인하고 스키마 일부를 조정하여 XML과 온톨로지 간의 연계 호환성을 높였다. 아울러 온톨로지 설계 과정에서의 주요 논의와 고민 지점을 함께 제시함으로써, 본 연구의 데이터 설계가 단순한 형식적 대응을 넘어 의미론적 구조화를 지향했음을 밝히고자 하였다. 3장에서는 온톨로지에 기반한 트리플 데이터를 편찬하기 위한 XML 데이터 파싱(Parsing) 작업을 다루었다. 파싱 작업에는 파이썬(Python)과 LLM(ChatGPT) 두 가지 디지털 도구를 활용하여, 기존의 규칙 기반 코딩 방식과 최근 시도되고 있는 LLM 기반 코딩 방식을 함께 적용하였다. 편찬된 트리플 데이터를 바탕으로 자료의 관계망을 시각화하는 지식 그래프를 일부 구현함으로써 경학 자료에 내재한 다층적 정보의 재현 가능성을 탐색하였다. 또 몇 가지 연구 질문을 반영한 사이퍼 쿼리(Cypher Query)를 생성하여 시맨틱 데이터의 연구 활용 방안과 자료 간 비교 연구로의 확장 가능성을 실험적으로 검토하였다.

2. 『상서고훈』, 『상서고주』 온톨로지 디자인

온톨로지(Ontology)라는 용어는 본래 철학에서 ‘존재론’을 가리키는 개념으로, 존재의 본질과 그 구조를 탐구하는 학문이다. 온톨로지는 여러 분야에 확장되어 다양한 의미로 사용되고 있는데, 정보과학 분야에서의 온톨로지는 ‘정보화의 대상이 되는 세계를 전자적으로 표현할 수 있도록 구성된 데이터 기술 체계’를 의미한다.² 다시 말해, 디지털 환경에서의 온톨로지는 개별 요소와 요소 간의 관계, 그리고 이들의 속성이 유기적으로 결합한 데이터의 총체를 정의하는 개념으로 활용된다.

디지털 인문학 영역에서 온톨로지 기반의 연구는 꾸준히 진행되어 왔으며, 한문학 분야에서도 역시 장르별로 구체적인 연구 성과가 제출되었다.³ 이러한 연구들은 인문지식에 내포된 복합적 의미망을 디지털 공간에서 구현하는 과정에서 온톨로지 설계가 주요한 방법론으로 기능하고 있음을 보여준다.

정보처리 방식의 측면에서 볼 때 온톨로지 설계는 기호주의(Symbolism)에 기반하고 있다. 기호주의는 세계에 존재하는 수많은 지식을 인간이 기호로 변환하여 컴퓨터에 입력하고, 정의된 규칙 체계와 논리적 구조에 기반해 이를 단계적으로 계산하고 추론해 나가는 방식이다. 반면 연결주의(Connectionism)는 인공 신경망이 다수의 외부 입력을 받아 경험적으로 학습하고, 이를 바탕으로 특정한 결과를 도출할 수 있다고 본다. 인간이 세계의

2 김현·안승준·유인태(2018), pp. 23.

3 대표적 시맨틱 데이터 모델링 사례로는 누정기(박순, 2017), 고문서(김현·안승준·유인태, 2018), 생활일기(유인태, 2019), 야담(양승복·유인태, 2022), 여항문학(김지선·유인태, 2022), 시화(한희연·지영원·호정·유인태, 2022), 한시(지영원·최진경, 2024), 소설(이길환, 2024), 경학(변은미·이동학·유인태, 2024), 과시자료(이길환·변은미·유인태, 2024), 관원록(김바로, 2023), 그 외 한국문학(장문석·유인태, 2021), 불교학(김바로, 2018), 불가 유람기(장지훈, 2025) 등을 들 수 있다. 본 목록은 이길환·변은미·유인태(2024, pp. 70-71.)의 논문에서 제시된 시맨틱 데이터 모델링 사례 분류를 바탕으로 하였으며, 여기에 이후의 연구 성과를 보완하여 정리한 것이다.

모든 지식을 규칙에 근거해 일일이 입력하는 것이 불가능하다는 점에서 기호주의는 한계를 지적받았고, 딥러닝의 부상과 함께 정보처리의 헤게모니는 연결주의로 이동하게 되었다.⁴

이러한 흐름 속에서 디지털 인문학 분야에서도 연결주의 철학에 입각한 딥러닝 기반 연구가 활발히 전개되고 있다. 하지만 한문 자료, 특히 본 연구에서 다루는 경학 자료의 경우 여전히 규칙 기반의 온톨로지 설계가 유효하다. 경학 자료는 경문과 주석 간에 층위가 존재하고, 주석 간에도 상호 맥락이 형성되어 있다. 경문에 대한 주석은 자구(字句)를 풀이하는 훈고(訓詁), 우주론·심성론 등으로 개념을 확장하여 해석하는 의리(義理), 경문을 여타 문헌을 통해 실증적으로 입증하고자 하는 고증(考證)의 성격을 지니며, 주석 내부의 지식 구조 역시 여러 해석 간의 복합적인 연계 속에서 형성된다. 경학 자료의 지식 구성 방식이 다양한 의미 기반의 관계를 통해 조직된다는 점에서, 기호주의적 접근에 기반한 온톨로지 설계가 여전히 유효하며 적합하다고 할 수 있다.

개별 주석 간의 관계를 설정하고 그것이 경문과 어떤 양상으로 연결되는지를 고찰하는 이러한 방식의 연구는, 본문의 이해에 기여할 뿐 아니라 주석가가 경전을 이해하는 총체적인 사고구조를 드러낸다는 점에서도 강점을 지닌다. 다시 말해, 아날로그적 연구 환경에서는 체계적으로 조직하기 어려웠던 경학 자료의 다층적 관계망을 디지털 환경에서 유연하고 풍부하게 재구성할 수 있는 가능성을 열어준다.

경학 자료 가운데 특히 『상서』는 『금문상서』와 『고문상서』 논쟁, 다양한 이문(異文)의 발생, 일실(逸失)된 경문의 복원 시도 등 경학사적 쟁점이 자구마다 함축적으로 담겨 있는 텍스트이다. 이러한 점에서 연구자의 해석적 관점을 바탕으로 온톨로지를 설계한다면, 『상서』 학술사의 전반적인 흐름과 경전을 복원하고 해석하고자 하는 주석가의 악전고투를 디지털 환경에서 구현할 수 있을 것으로 기대된다.

4 박승일(2024), pp. 51-52.

2.1. 온톨로지 디자인의 기본적 구조

경학 자료를 디지털 환경에서 구현하기 위해 고려한 점은 원전의 형태를 얼마나 충실히 반영할 수 있을지, 문면에 나타나지 않는 함의를 어떻게 효과적으로 드러낼 수 있을지였다. 특히 본 연구에서는 『상서고훈』과 『상서고주』를 연계한 온톨로지를 설계함으로써 두 저작의 공통점을 조망하면서도 각 저작만의 고유한 특징을 뚜렷하게 예각화하고자 하였다.

경학 자료를 연구할 때는 자료의 기본적인 구조와 내용을 함께 살펴볼 필요가 있다. 경학 자료의 특징은 경전 원문과 각 주석이 명확한 위계질서를 바탕으로 제시된다는 점이다. 경전 원문은 분절 단위에 따라서, 권, 편, 장, 절, 구 등으로 구성되며, 다시 각 단위별로 주석이 존재한다. 주석은 경학자가 원문을 이해한 방식을 보여준다고 할 수 있다. 인용한 자료와 구절, 다시 그에 대한 저작의 견해, 논평 및 기타 언급은 경학자의 주해(注解) 과정을 보여주는 지표이다. 이에 본 연구에서는 원문의 구성 항목으로서, 권, 편, 장, 절, 구를, 주석의 구성 항목으로 인용, 구절, 견해, 논평, 언급을 설정하였다.

또한 각 주석이 어떤 학술사적 맥락에서 서술되었고 주석 텍스트에 내포된 문맥은 무엇인지를 파악하기 위해 그 근거가 되는 출처인 문헌과 인물을 문맥 정보로 설계하였다.

한편 자료에는 직접적으로 나타나 있지 않으나 연구자가 학술사적 의의를 판단하기 위해 개입하는, 즉 연구자의 해석이 반영되어야 하는 부분이 있다. 우선 상서학사에서 가장 주요한 논쟁점인 『금문상서』, 『고문상서』와 관련된 내용을 담기 위해 ‘금고문’ 항목을 설정하였고, 주석을 서술할 때 제시된 문헌이 아닌 2차 자료를 보고 인용한 경우가 있는데, 이는 학자의 자료 운용 방식을 면밀하게 분석할 수 있는 단서가 되므로 ‘재인용’ 항목을 설계하였다.

[그림 1]은 『상서고훈』, 『상서고주』의 온톨로지 설계를 위한 구상도로



[그림 1] 『상서고훈』, 『상서고주』 온톨로지 디자인의 개념적 구조

서, 경학 자료를 디지털 환경에서 구조화하고 연결하기 위한 개념적 체계를 시각화한 것이다. 우선 경학 자료를 구성하는 형식 정보로서 원전 정보와 주석 정보를 설정하였으며, 주석 정보를 구성하는 핵심 문맥 정보로 인물과 문헌을 주요 개체로 삼아 주석의 지식적 출처를 명확히 하고자 하였다. 이와 함께 연구자의 해석이 반영된 정보를 별도의 해석 요소로 구분하고, 여기에 ‘금고문’과 ‘재인용’ 항목을 두었다.

이상의 네 가지 주요 정보는 기존 XML 스키마의 핵심 구성 축과 구조적으로 대응하며 그 설계 원리와의 일관된 맥락을 공유하는데, 이는 XML 설계와 온톨로지 디자인이 상호 유기적인 관계 속에서 각각의 구조적 원리 및 목적을 반영하며 통합·교차적으로 설계되었기 때문이다. 따라서 전체적인 구조와 설계 원리의 측면에서는 XML 스키마와 온톨로지가 긴밀하게 대응한다. 다만 세부적인 연계 방식에서는 양자 간에 차이가 존재하는데, 이러한 차이점은 다음 절에서 구체적으로 서술하겠다.

2.2. XML 스키마와 온톨로지의 연계 방식

시맨틱 데이터 모델의 기반이 되는 온톨로지는 정보 간의 의미적 관계를 명시적으로 기술하기 위한 체계로서, XML 스키마에서 정의된 요소

들은 온톨로지와 의 연계를 통해 그 의미와 맥락을 부여받는다. XML 스키마의 요소(Element)와 속성(Attribute)은 온톨로지에서 클래스(Class), 속성(Attribute), 관계(Relation) 등으로 반영되는데, 구체적인 연계 방식은 [표 1]에 정리하였다.

[표 1] XML 스키마의 온톨로지 반영 맥락

대분류	중분류	소분류	Element	온톨로지 반영 맥락	
형식 요소	원전 요소	서	Section	[Class] Section	
		권	Volume	[Class] Volume	
		편	Chapter	[Class] Chapter	
		장	Paragraph	[Class] Paragraph	
		구	Line / Unit	[Class] Line	
	주석 요소	형태	주석	Commentary	id, name 값에 언더바(_) 표기
			소주	SCommentary	[Attribute] form
			난외주	MCommentary	[Attribute] form
		영역	훈고	Glossary	[Attribute] area
			교감	Collation	[Attribute] area
			분류표지	Marker	[Attribute] area
		방식	인용	Citation	[Class] Citation
			견해	Opin	[Class] Opinion
			언급	Ment	[Class] Mention
			논평	Remark	[Class] Remark
		유형	출처	Refer	[Relation] A mentions B
			구절	Text	[Class] Text
		문맥 요소	문헌	Book	[Class] Book
			편목	Part	[Relation Attribute] part
비문	Epigraph		[Class] Book		
인물	Person		[Class] Person		
해석 요소	금고문	NewOldText	[Class] NewOldText		
	재인용	Recite	[Class] Recite		
	연구자 교감	Edit	-		

XML 스키마의 원전 요소를 구성하는 개별 분절 항목은 모두 온톨로지의 클래스로 생성하였다. 원전 클래스를 생성한 뒤에 분절 단위를 속성값으로 처리하는 방식도 고려할 수 있었으나, 분절 요소들이 각각의 고유한 주석 정보와 의미를 지니고 있어서 개별 단위 모두를 독립된 클래스로 설계하였다.

주석 요소의 경우 형태 범주에 속하는 <Commentary>, <SCommentary>, <MCommentary> 가운데 <Commentary>는 원문과 주석을 구분하는 역할을 하면서 모든 주석 요소를 포괄하는 상위 태그이다. 따라서 온톨로지에서는 별도의 클래스로 생성하지 않고, id와 name 값에 언더바(_) 표기를 사용하여 원문과 구분되도록 하였다. 이 표기는 주석임을 명시하는 동시에 주석이 귀속되는 원문을 식별할 수 있게 한다. 한편 <SCommentary>와 <MCommentary>는 인용(Citation) 클래스의 속성값으로 반영되어, 인용 간 관계를 파악하고 관계 데이터를 생성하기 위한 지표로 활용된다.

주석 요소 가운데 온톨로지에서 클래스로 범주화된 개체는 인용(Citation), 견해(Opinion), 논평(Remark), 언급(Mention), 구절(Text)이며, 구절을 제외한 네 항목은 정확자의 주석 활동 방식을 반영한다. 인용과 견해 클래스는 원전 정보와 직접 연결되며, 인용-견해, 인용-논평, 인용-언급 등의 의미적 연계를 통해 자료 내 지식의 구조화 양상을 드러낸다. 구절 클래스는 인용된 주석의 텍스트 단위를 담기 위한 틀로서, 외부적으로는 주석 정보와 문맥 정보를 연결하는 매개 역할을 수행한다. 이때 출처를 의미하는 <Refer> 요소는 온톨로지에서는 'A mentions B' 관계 데이터로 변환되어 문헌-구절, 인물-구절 간의 연결을 가능하게 한다.

문맥 요소 <Book>, <Part>, <Epigraph>는 모두 문헌의 범주에 속하므로 온톨로지에서는 문헌(Book) 클래스로 포섭되는데, 각각의 기능적 역할에는 차이가 있다. 이 가운데 <Part>는 문헌 클래스와 구절 클래스 간의 관계를 매개하는 속성값으로 사용되며, 특정 구절이 문헌 내에서 어느 부분에 속하는지를 명시한다는 점에서 문헌의 내부 구조를 기술하는 계층적 정보로서

가능하다. 따라서 <Part>는 문헌-구절 관계의 속성값으로 기술되며, 구절의 위치와 해석 맥락을 파악하는 중요한 단서를 제공한다. <Book>, <Part>, <Epigraph> 요소를 문헌 클래스의 범주로 통합하는 과정에서 <Book>에는 해당 편명을, <Part>에는 원자료명을 XML 속성으로 명시하여 각 요소가 지칭하는 문헌 정보의 식별성을 강화하고 데이터 추출의 효용성을 높일 수 있도록 XML 스키마를 보완하였다. <Epigraph>의 경우 문헌 클래스의 속성 정보에 별도의 분류 체계를 마련하여 해당 요소를 식별할 수 있도록 하였다. <Person>은 인물(Person) 클래스로 반영되었다.

해석 요소 <NewOldText>와 <Recite>는 온톨로지에서 각각 금고문(NewOldText)과 재인용(Recite) 클래스로 설계하여 해당 개체와 연결하였다. 한편 <Edit>는 연구자가 교감한 내용을 표시하는 태그이므로 온톨로지에서는 별도의 데이터로 생성하지 않았다.

2.3. 『상서고훈』, 『상서고주』 연계 온톨로지 디자인

2.3.1. 클래스·속성 데이터 설계 내용과 그 근거

① 형식 정보(Textual Information) 1: 원전 정보(Original Information)

원전 정보는 주석의 대상인 경전 원문을 담기 위해 설계된 클래스로 주석 정보와 직결된다. 경전 원문은 편(篇), 장(章), 절(節) 등 다양한 분절 단위를 가지며, 분절 기준에는 개별 학자의 경학관이 반영되는 경우가 많다. 또한 각 분절 단위에는 해당 주석이 대응되므로, 단위별로 모두 개별 클래스를 설계하고 자료별 분절 기준에 따라 생성하고 연결하였다.

『상서』는 서, 권, 편, 장, 구의 분절 단위를 기본으로 하되, 자료별 분절 기준에는 부분적인 차이가 있다. 『상서고훈』의 경우 ‘서’의 구분 없이 ‘권’이 최상위 분절 단위로 설정되었으며, 『상서고주』는 ‘장’ 구분을 따로 두지 않았다. 또한 두 자료는 『금문상서』(今文尙書)만을 대상으로 한 저술이므로 「순전」(舜典), 「대우모」(大禹謨) 등 『위고문상서』(僞古文尙書)에 해당하는 편들

[표 2] 『상서고훈』, 『상서고주』 원전 정보 클래스·속성 설계와 예시

Class	Attribute	『상서고훈』 예시	『상서고주』 예시
Section	id	-	S1
	name	-	우하서
	content	-	虞夏書
Volume	id	V1	V1
	name	권1	권1
	content	卷一	卷一
Chapter	id	C01	C01
	name	요전	요전
	content	堯典	堯典
Paragraph	id	C01P01	-
	name	요전1장	-
	content	曰若稽古 …… 格于上下	-
Line	id	C01P01L01	C01L001
	name	요전1장1구	요전1구
	content	曰若稽古	曰若稽古

은 제외되었다.

② 형식 정보(Textual Information) 2: 주석 정보(Annotative Information)

주석 정보는 원전 정보에 직접 연계되는 정보로서, 경학자의 원문 대응 방식을 기준으로 클래스를 설계하였다. ‘인용’(Citation)은 경학자가 원문에 대한 3자의 해설을 인용함으로써 간접적으로 대응한 주석이고, ‘견해’(Opinion)는 경학자가 원문에 대해 자신의 의견을 개진하여 직접적으로 대응한 주석이다. 이 두 클래스는 주석 체계의 최상위 범주로서 원문과 직결되는 정보이다.

추가로 견해 클래스 내부에는 앞서 인용한 주석의 시비를 평가한 ‘논평’(Remark) 클래스와 인용 중간중간에 나타나는 ‘인용’에 대한 경학자의 추

가 견해를 담은 ‘언급’(Mention) 클래스를 함께 설계하였다. 논평은 견해 내부에 존재하면서 인용과 직결되며, 언급은 인용 사이에 배치되어 인용과 연결된다. 이 두 클래스는 모두 ‘인용’에 대한 경학자의 직접적 대응을 담고 있다는 점에서, 주석 내부 정보 간의 의미 연결성을 드러낸다.

또한 인용 주석의 본문에 해당하는 ‘구절’(Text) 클래스를 생성하여 문헌, 인물 등 출처에 해당하는 문맥 정보와의 연결을 고려하였다. 인용-구절, 구절-출처의 순차적 관계를 지식 그래프상에서 명확히 표현하기 위해 ‘구절’ 클래스를 중간 노드로 설정하였다.

◎ 인용(Citation) 클래스

[표 3] 『상서고훈』, 『상서고주』 인용(Citation) 클래스 속성 설계와 예시

Attribute	구분	설명	『상서고훈』 예시	『상서고주』 예시
id	식별자	원문 id+언더바+영역(G/C)+CI+순서	C01P01L07_GCI2	C01L007_CCI4
name	대표명	원문 name+언더바+영역명+순서+인용	요전1장7구_훈고2인용	요전7구_교감4인용
area	영역	훈고·교감·분류표지	훈고	교감
form	형태	기본·소주·난외주	소주	기본
format	형식	직접인용·간접인용	간접인용	직접인용
context	맥락	단순인용·긍정·부정·부분긍정 // 참조·근거·비판	긍정	단순인용
content	내용	인용 부분 전체 입력	蔡說 同	徐廣曰 駟古訓字 順也
original text	원문	간접인용의 경우 해당 원문을 찾아서 입력	格 至 上 天 下 地 也 言 其 德 之 盛 如 此	-

인용 클래스는 경학자가 원문에 대한 3자의 해설을 인용한 부분을 담은 틀로, 원문·견해·인용·논평·구절 클래스와 직결된다. id(식별자),

name(대표명), area(영역), form(형태), format(형식), context(맥락), content(내용) 총 7가지 속성을 설계하였다. 원문 클래스의 id와 name 값에 언더바(_)를 표기하여 주석 요소임을 구분한 다음 ‘영역(Glossary/Collation) + 인용(CI) + 순서’, ‘영역명 + 순서 + 인용’의 기술 논리로 인용 클래스의 id와 name 값을 입력하였다.

id와 name 값에 반영된 ‘영역’ 정보는 인용 클래스의 속성으로 별도로 설계하였다. 이는 각 경학 자료에 내재한 고유 분류 체계를 최대한 반영하기 위함이다. ‘훈고’와 ‘교감’ 등의 영역 정보를 원문과 인용 클래스를 연결하는 관계 데이터로 구성하는 방안도 검토하였으나, 인용과 원문 간의 관계를 단순히 인용 주석의 성격으로 환원하기보다는 인용자가 어떠한 해석적 맥락 속에서 해당 주석을 인용하였는지에 주목할 필요가 있었다. 이는 해당 데이터 모델이 어떤 주석을 인용하였고, 몇 번 인용하였는지에 대한 통계적 연구를 지향하는 것이 아니라 주석 행위의 복잡성과 입체성을 조망할 수 있는 구조를 구축하는 데 목적이 있기 때문이다.

form(형태)은 제시된 인용의 형태를 구분하는 속성으로, 인용 정보의 층위를 반영하기 위해 설계되었다. 소주(小注) 형태로 제시된 인용은 앞선 주석에 종속되어 그 내용을 해설하거나 보완하는 기능을 수행하는 경우가 많으며, 난외주(欄外注) 형태의 인용은 이후에 추가되었을 가능성이 높고 특정 의도나 관점이 반영된 사례가 많다. 이러한 형태 속성은 이후 인용 간 관계를 정의할 때 인용 정보들 사이의 종속적-보완적 구조를 명확히 하고, 주석 간의 맥락적 연결을 표현하는 데 활용된다.

인용 구절이 원문 내용을 그대로 제시한 경우를 직접인용으로, 원문을 요약하거나 변형하여 제시하거나 출처를 명시하지 않은 경우를 간접인용으로 구분하여 format(형식) 속성에 기술하였다. 이 속성 역시 경학자의 인용 방식을 면밀히 분석하기 위해 설계되었다. 예컨대 신작의 경우 경학 자료를 저술함에 있어 직접적인 견해 표출을 지양하는 서술방식을 취하였다. 이러한 서술방식의 특성상 원문을 재구성하여 제시한 간접인용이 오히려

려 신작의 해석과 견해를 일정 부분 반영한 주요 정보로 기능한다. 또한 교감을 목적으로 한 주석의 경우 대부분 간접인용의 형식을 취한다는 특징을 지닌다.

인용 맥락을 반영하기 위해 context(맥락) 속성을 설계하였으며, 이 속성은 ‘인용-원문’, ‘견해 내 인용-견해’, ‘인용 내 인용-인용’ 간 관계 연결의 기반이 된다. 인용 맥락은 ‘단순 인용·긍정·부정·부분 긍정’과 ‘참조·근거·비판’으로 다시 구분되는데, 전자는 ‘인용-원문’ 간의 관계를 규정하는 기준이며 후자는 ‘견해 내 인용-견해’, ‘인용 내 인용-인용’, 즉 주석자가 자신의 해석 안에 또 다른 인용을 배치한 경우를 설명하는 기준으로 설정하였다.

인용-원문 간의 관계는 저자의 인용 맥락을 기준으로 ‘단순 인용·긍정·부정·부분 긍정’의 네 가지 유형으로 세분화하였다. 이 가운데 단순 인용은 인용자가 해당 주장에 대해 특정한 평가적 입장을 취하지 않고 객관적인 정보 전달을 목적으로 인용한 경우를 의미한다. 나머지 세 유형은 인용자의 평가적 입장이 드러나는 사례로, 긍정 인용은 인용자가 해당 주장에 동의하거나 지지하는 맥락에서 인용한 경우, 부정 인용은 반박하거나 비판하는 입장에서 인용한 경우를 가리킨다. 부분 긍정 인용은 인용한 주장의 일부 요소에 대해서는 동의하지만 다른 부분에 대해서는 유보적인 태도를 취하는 경우를 말한다. 다산은 대부분의 인용에 대해 자신의 견해를 명시함으로써 인용 맥락을 비교적 분명히 드러낸 반면, 석천은 별도의 해설 없이 인용 주석을 나열한 경우가 많다. 따라서 『상서고주』에는 단순 인용의 사례가 많이 확인된다.

견해 내 인용-견해, 인용 내 인용-인용 간의 관계는 인용자가 자신의 견해를 제시하는 과정에서 제3자의 주장을 다시 인용하는 경우를 가리킨다. 이러한 관계는 앞서 설명한 인용-원문 간 관계와 유사한 기준으로 인용 맥락을 구분할 수 있으나, 텍스트 내 계층 구조에서 차이를 보인다. 즉, 인용-원문 관계는 원문과 주석 간의 층위 차이를 전제로 하는 반면, 견해 내 인용-견해 및 인용 내 인용-인용 관계는 모두 주석 내부, 즉 주석과 주석

간의 층위에서 형성된다는 점에서 구별된다. 이에 따라 해당 관계는 인용자가 자신의 견해를 뒷받침하기 위한 근거 인용, 반박을 목적으로 한 반박 인용, 단순한 정보 제공이나 연계적 언급을 위한 참조 인용의 세 가지 맥락으로 별도로 설계하였다.

◎ 견해(Opinion) 클래스

[표 4] 『상서고훈』, 『상서고주』 견해(Opinion) 클래스·속성 설계와 예시

Attribute	구분	설명	『상서고훈』 예시	『상서고주』 예시
id	식별자	원문 id+'_' +영역+O+순서	C01P01L01_GO1	C01L017_CO1
name	대표명	원문 name+'_' +영역+순서+견해	요전1장1구_훈고1견해	요전17구_교감1견해
area	영역	훈고·교감·분류표지	훈고	교감
content	내용	견해 부분 전체 입력	鏞案 …	緯按 …

견해 클래스는 경학자가 원문에 대응하여 자신의 견해를 개진한 내용을 담는 틀로, 원문·인용·논평 클래스와 직결된다. id(식별자), name(대표명), area(영역), content(내용) 총 네 가지 속성을 설계하였다. 전체적인 속성 구성 및 id와 name 값의 기술 논리는 인용 클래스와 동일하며, 'CI'는 'O'로, '인용'은 '견해'로 치환하여 적용하였다.

다산은 '용안'(鏞案)을 견해의 기본 표지로 사용하였으며, 이 외에도 '변왈'(辨曰), '정왈'(訂曰), '의왈'(議曰) 등 다양한 표지를 통해 자신의 견해를 입체적이고 적극적으로 드러냈다. 석천 또한 '작안'(緯按) 등의 표지를 사용하여 견해를 제시하였으나, 신작의 견해는 대부분 인용 주석의 나열로 구성되어 있다는 점에서 다산과 차이를 보인다. 다산의 견해가 자신의 논점과 해석을 분명하게 드러내는 주체적이고 논쟁적인 서술 방식으로 구성된 반면, 석천은 견해를 명시적으로 개진하기보다는 기존 주석의 다양한 인용을 통해 암묵적이고 우회적으로 자신의 입장을 드러내는 방식을 취하였다.

◎ 논평(Remark) 클래스

[표 5] 『상서고훈』, 『상서고주』 논평(Remark) 클래스 속성 설계와 예시

Attribute	구분	설명	『상서고훈』 예시	『상서고주』 예시
id	식별자	견해 id+RM+순서	C01P01L01_GO1RM1	-
name	대표명	견해 name+순서+논평	요진1장1구_훈고1견해1논평	-
area	영역	훈고·교감·분류표지	훈고	-
object	대상	논평 내 인용 지칭 부분	蔡說	-
content	내용	논평 부분 전체 입력	蔡說是也……………詞例 至臯陶謨而不通也	-

논평 클래스는 견해 내부에서 앞서 제시한 인용 주석에 대해 논평한 내용을 담는 틀로, 견해·인용 클래스와 직결된다. id(식별자), name(대표명), area(영역), 대상(object), 내용(content) 총 5가지 속성을 설계하였다.

논평은 견해에 종속된 정보이므로, id와 name 값에는 견해 클래스의 값을 기반으로 각각 'RM'과 '논평'을 덧붙이는 방식으로 기술하였다. 또한 논평은 특정 인용에 대한 직접적인 대응 주석이므로, 그 대상이 되는 인용 정보를 명확히 연결하기 위한 속성으로 object(대상)를 별도로 설계하였다. 논평 클래스는 인용자의 평가적 의도가 반영된 인용 맥락을 구분하는 기준으로 기능하며, 본 연구의 대상 자료 가운데 『상서고훈』에서만 확인된다.

◎ 언급(Mention) 클래스

[표 6] 『상서고훈』, 『상서고주』 언급(Mention) 클래스·속성 설계와 예시

Attribute	구분	설명	『상서고훈』 예시	『상서고주』 예시
id	식별자	인용 id+MT+순서	C01P03L02_GCI5MT1	C01L006_CCI3MT1
name	대표명	인용 name+순서+언급	요전3장2구_훈고5인 용1언급	요전6구_교감3인용 1언급
area	영역	훈고·교감·분류표지	훈고	교감
content	내용	언급 부분 전체 입력	評曰 夏氣不可云高明	並以光爲廣也

언급 클래스는 인용 중간중간에 간헐적으로 보이는 경학자의 언급을 담는 틀로, 인용 클래스와 직결된다. id(식별자), name(대표명), area(영역), content(내용) 총 네 가지 속성을 설계하였다.

언급은 인용에 종속된 정보이므로, id와 name 값에는 인용 클래스의 값을 기반으로 각각 ‘MT’와 ‘언급’을 덧붙이는 방식으로 기술하였다.

◎ 구절(Text) 클래스

[표 7] 『상서고훈』, 『상서고주』 구절(Text) 클래스·속성 설계와 예시

Attribute	구분	설명	『상서고훈』 예시	『상서고주』 예시
id	식별자	인용 id+T	C01P01L01_GCI3T	C01L005_CQ1T
name	대표명	인용 name+구절	요전1장1구_훈고3인용 구절	요전5구_교감1인용 구절
content	내용	구절 부분 전체 입력	曰若者 發語辭 史臣將敘 堯事 故先言考古之帝堯	合於堯之克攘
reference	출처	구절과 함께 제시된 출처를 입력	蔡	漢藝文志
reference type	출처 유형	제시된 출처의 유형 을 입력	Person	Book, Part
reference target	출처 대상	제시된 출처의 온전 한 명칭을 입력	蔡沈	漢書 藝文志

구절 클래스는 인용 구절을 담는 틀로, 인용·문헌·인물 클래스와 직결된다. id(식별자), name(대표명), content(내용), reference(출처), reference type(출처 유형) reference target(출처 대상) 총 6가지 속성을 설계하였다. 이 가운데 reference와 관련한 3가지 속성을 별도로 설계한 이유는, 구절 클래스가 주석 정보와 문맥 정보를 연결하는 중간 노드로서 기능하기 때문이다. 이러한 속성들은 인용된 정보의 출처를 인물, 문헌, 편명 등 다양한 수준에서 정밀하게 식별할 수 있도록 하여 출처 기반의 의미적 연결 관계를 구성하는 기반이 된다.

③ 문맥 정보(Contextual Information)

문맥 정보는 문맥 내에서 객관적으로 식별 가능한 정보를 가리킨다. XML 스키마에서는 ‘문헌’, ‘파트’, ‘인물’, ‘비문’을 문맥 요소로 설계하였는데, 온톨로지에서는 ‘문헌’, ‘파트’, ‘비문’을 모두 문헌(Book) 클래스로 포섭하고, ‘인물’ 요소는 그대로 인물(Person) 클래스로 유지하였다. 이 두 항목은 자료 내에서 주석의 근거가 되는 출처로 기능하며, 개별 경학자의 자료 운용 방식을 분석하기 위한 토대가 된다.

앞서 소개한 형식 정보와는 달리 문맥 정보에 해당하는 문헌, 인물 클래스의 개체는 주석에서 참조하는 고정된 실체를 지칭하는 정보로서, 반복적으로 사용되는 고유 용어이자 별도의 메타데이터에 기반하여 일관되게 관리되어야 할 개념이다. 따라서 해당 클래스의 개체들은 별도의 글로서리(glossary) 항목으로 정리되어야 하며, 용어의 표준화와 통합 관리가 필요하다. 아울러 클래스 내 개체 간의 관계 역시 온톨로지 설계의 대상이 될 수 있음을 고려하고 있다.

◎ 문헌(Book) 클래스

[표 8] 문헌(Book) 클래스·속성 설계와 예시

Attribute	구분	설명	예시	
id	식별자	B+언더바+문헌 이니셜+구별 번호(입력순)	B_SG01	B_SG02
korname	한글명	서명의 한글명 입력	사기	시경
chiname	한자명	서명의 한자명 입력	史記	詩經
classification	분류	經·史·子·集·碑文 분류	史	經

문헌 클래스는 문헌 정보를 담는 틀로, 구절·인물 클래스와 직결된다. 이 클래스에는 인용 출처에 해당하는 문헌, 연구 대상으로 삼은 경서 주석서, 단순 언급된 문헌까지 모두 포함된다. id(식별자), korname(한글명), chiname(한자명), classification(분류) 총 네 가지 속성을 설계하였다. korname과 chiname에는 문헌의 공식 서명을 각각 한글과 한자로 기술하였으며, classification(분류)은 경(經)·사(史)·자(子)·집(集), 비문(碑文) 등의 문헌 분류를 위한 속성으로, 향후 세분화 및 생성 기준에 대한 추가 논의를 고려하고 있다.

◎ 인물(Person) 클래스

[표 9] 인물(Person) 클래스·속성 설계와 예시

Attribute	구분	설명	예시
id	식별자	P+언더바+인명 이니셜+구별 번호(입력순)	P_JH01
korname	한글명	인물의 한글명 입력	정현
chiname	한자명	인물의 한자명 입력	鄭玄
birthyear	생년	인물의 생년을 입력	127
deathyear	몰년	인물의 몰년을 입력	200
nationality	국적	한국·중국·일본 등 구분	중국

인물 클래스는 인물을 담는 틀로, 구절·문헌 클래스와 직결된다. id(식별자), korname(한글명), chiname(한자명), birthyear(생년), deathyear(몰년), nationality(국적) 총 6가지 속성을 설계하였다.

④ 해석 정보(Interpretative Information)

연구자의 해석적 개입이 반영된 해석 정보로 ‘재인용’(Recite) 클래스와 ‘금고문’(NewOldText) 클래스를 설계하였다. 경학 데이터 모델을 통해 개별 경학자의 자료 운용 방식을 면밀히 분석하기 위해서는 해당 인용이 원자료에서 직접 이루어진 것인지, 아니면 제3의 자료를 통해 간접적으로 이루어진 것인지를 구분할 필요가 있다. 이를 위해 2차 자료를 통해 재인용된 것으로 추정되는 사례를 담기 위해 재인용 클래스를 별도로 설계하였다.

금고문 클래스는 『상서』, 『춘추』 등의 일부 경학 자료에서 논의되는 금고문 관련 담론을 담기 위한 틀로, 특정 경전의 해석 과정에서 주요한 맥락을 지니는 정보를 반영하기 위해 설계되었다. 한편 기존 XML 스키마에 포함되어 있던 〈Edit〉 요소는 온톨로지상에서 별도의 의미적 맥락을 형성하지 않는다고 판단되어 제외하였다.

◎ 재인용(Recite) 클래스

[표 10] 재인용(Recite) 클래스·속성 설계와 예시

Attribute	구분	설명	『상서고훈』 예시	『상서고주』 예시
id	식별자	RE+순서	RE003	RE001
name	대표명	재인용+순서	재인용003	재인용001
content	내용	재인용 부분 전체 입력	書大傳 舜巡四嶽 祀泰山 霍山 皆奏義伯之樂 華山 弘山 奏和伯之樂	考靈耀 放勛欽明文 塞晏晏 鄭彼注 寬容 覆載謂之晏
book	서명	재인용 추정 자료의 서명	資治通鑑前編	後漢書
original text	원문	재인용 추정 자료의 원문 입력	履祥按 尙書大傳 舜巡四嶽 祀泰山霍山 皆奏義伯之樂 華山弘山 奏和伯之樂	尙書考靈耀曰 放勛欽 明文思晏晏 鄭彼注 寬容覆載謂之晏

재인용 클래스는 2차 자료에서 재인용한 사례로 추정되는 경우를 담은 틀로, 인용 클래스와 직결된다. id(식별자), korname(한글명), content(내용), book(서명), original text(원문) 총 5가지 속성을 설계하였다. 연구자의 추정이 반영된 정보이므로 객관성을 담보하기는 어렵지만, 정보가 누적되면 문헌 활용 방식의 구체적 경향을 추적하는 데 중요한 단서를 제공할 수 있다.

◎ 금고문(NewOldText) 클래스

[표 11] 금고문(NewOldText) 클래스 속성 설계와 예시

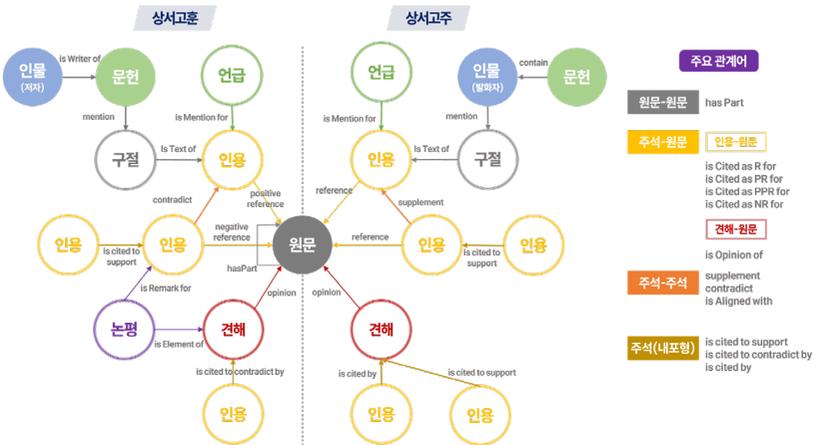
Attribute	구분	설명	『상서고훈』 예시	『상서고주』 예시
id	식별자	NO+순서	NO001	NO008
name	대표명	금고문+순서	금고문001	금고문008
content	내용	금고문 부분 전체 입력	僞註順考古道之解 遠自西京其源遠矣 僞者深信其義 既於此註之如此 又撰周官之經儼稱唐虞稽古 建官維百 則堯舜二帝 遂爲稽古之主 不可辭矣	今文及馬鄭王本 堯舜典合爲一 梅賾所上孔傳古文 猶闕舜典 乃分慎徽五典以下 爲舜典之初

금고문 클래스는 고문상서·금문상서 관련 내용을 담은 틀로, 인용·견해·논평·언급 클래스와 직결된다. id(식별자), korname(한글명), content(내용) 총 세 가지 속성을 설계하였다. 금고문 논의의 주요 쟁점과 전개 양상을 독립적으로 살펴보기 위해 이 클래스를 구성하였다.

2.3.2. 관계 설계와 그 연계적 의미

관계 데이터 설계에는 경학 자료를 구성하는 정보들을 의미적으로 연결하기 위한 고민이 반영되어 있다. 원문 간의 연결은 경학자가 설정한 분절 구조와 그 의도를 드러내며, 주석-원문 간의 연결은 경학자의 주석 활동 방식과 해석 접근법을 보여준다. 또 인용을 구성하는 여러 정보 간의 연결은 다양한 인용 제시 방식을 포괄함으로써 경학자의 문헌고증 방식을 유형

화하는 단서를 제공한다. 견해 내에서 조직된 여러 정보 간의 관계는 경학자의 논거에 활용된 내적 논리와 외적 참조의 관계를 분석하는 토대가 된다. 그리고 인용 간의 관계는 다양한 인용의 교차적 구성을 조직화함으로써 해석의 계보성과 경학자의 자료 배치 의도 등을 심층적으로 파악할 수 있게 한다. 해석 정보의 연결은 문헌의 활용 맥락 및 특정 쟁점에 대한 사고 과정을 추적할 수 있게 함으로써 질적 연구를 위한 토대를 제공한다.



[그림 2] 『상서고훈』, 『상서고주』 온톨로지 디자인

『상서고훈』, 『상서고주』의 온톨로지 디자인은 『상서』 원문을 중심으로 각 자료의 주석 정보인 인용과 견해가 연계되어 확장되는 구조를 취하고 있다. 원문-원문, 주석-원문, 주석-주석 간의 층위는 직접적·간접적 대응 관계에 따라 관계어를 구분하고, 전치사와 서술형 표현을 활용하여 이를 명확히 반영하고자 하였다. 이러한 기준에 따라 총 일곱 가지 유형, 21건의 관계 데이터를 생성하였다.

[표 12] 『상서고훈』 『상서고주』 관계(Relation) 설계

유형	Source (A)	Target (B)	Relation	설명	
원문-원문	원문	원문	hasPart	A는 B를 부분으로 가지다	
주석-원문	인용	원문	is Cited as □ for	reference(R)	단순 인용
				positive reference(PR)	긍정 인용
				negative reference(NR)	부정 인용
				partially positive reference(PPR)	부분 긍정 인용
	견해	원문	is Opinion of	A는 B에 대한 견해이다	
주석-주석 (기본형)	인용	인용	supplement	A는 B를 보충하다	
			contradict	A는 B를 반박하다	
			is Aligned with	A는 B와 같은 입장이다	
주석-주석 (내포형)	인용	인용	is Cited by	참조 인용	
			is Cited to support	근거 인용	
	견해	인용	is Cited to contradict by	반박 인용	
주석-주석 (종속형)	논평	견해	is Element of	A는 B의 요소이다	
	논평	인용	is Remark for	A는 B의 논평이다	
	언급	인용	is Mention for	A는 B에 대한 언급이다	
인용 내	구절	인용	is Text of	A는 B에 대한 구절이다	
	문헌	구절	mention	A는 B를 말하다	
	인물	구절			
	인물	문헌	is Writer of	A는 B의 저자이다	
	문헌	인물	contain	A는 B를 포함하다	
주석-해석	인용	재인용	re-cite	A는 B를 재인용하다	
	인용	금고문	include	A는 B를 포함하다	
	견해				
	논평				
	언급				

◎ 원문-원문 관계

원문의 경우 자료별 분절 단위에 따라 클래스를 생성하고, 각 단위는 “A hasPart B(A는 B를 포함하다)”의 관계어를 통해 계층적으로 연결하였다. 원문 간의 관계는 자료별 분절 구조를 반영한다. 『상서고훈』은 ‘서’를 최상위 분절 단위로 설정하지 않았으며, 『상서고주』는 별도의 ‘장’ 구분 없이 ‘편’ 하위에 바로 ‘구’를 배치하였다는 구조적 차이를 보인다. 또한 두 학자 모두 『고문상서』가 위작임을 확신하였기 때문에 『금문상서』에 해당하는 편만을 자료의 연구 대상으로 삼았다.

◎ 주석-원문 관계

주석과 원문 간의 관계는 경학자의 주석 활동을 대별하는 것으로 다시 인용-원문, 견해-원문 관계로 구분된다. 두 관계는 경학자의 주석 활동이 3자의 해설을 간접적으로 인용하는 방식과 본인의 의견을 직접적으로 개진하는 방식으로 나뉜다는 차이를 반영한 것이다. 이에 따라 두 관계의 관계어를 서술형 구조와 전치사 구성을 달리하여 설계하였으며, 이를 통해 인용을 매개로 한 간접적 해석인지, 견해를 통한 직접적 해석인지의 구분이 명확히 드러나도록 하였다.

인용-원문 관계는 해당 주석이 인용된 맥락을 반영하여 관계어로 설정하였으며, 인용 맥락은 단순 인용·긍정·부정·부분 긍정의 네 가지 유형으로 구분하였다. 단순 인용은 인용자가 해당 주장에 대해 명확한 입장을 취하지 않고 정보를 전달하기 위한 목적으로 인용하는 경우를 가리킨다. 이에 반해 나머지 세 가지 맥락은 인용자의 평가적 판단이 명확하게 드러나는 경우를 말한다. 긍정 인용은 경학자가 해당 주장에 동의하거나 지지하는 입장에서 인용한 경우이며, 부정 인용은 해당 주장을 비판하거나 반박하는 맥락에서 인용한 경우이다. 부분 긍정 인용은 인용한 주장에 대해 일부에 대해서는 찬성하거나 수용하는 입장을 보이지만, 다른 부분에 대해서는 비관적이거나 유보적인 태도를 보이는 경우에 해당한다.

다산은 각 주석의 인용 맥락을 대부분 견해를 통해 명확히 밝힌 반면, 석천은 주로 인용 주석만을 나열하고 이에 대한 추가 설명을 덧붙이지 않은 경우가 많다. 따라서 『상서고주』의 인용-원문 간 관계는 대부분 “A is Cited as reference for B(단순 인용)”의 형태를 보이고, 『상서고훈』은 긍정·부정·부분 긍정 등 다양한 인용 맥락이 확인된다.

인용자의 인용 의도를 판단할 때 연구자의 주관적 판단이 개입될 여지를 최소화하기 위해, 기본적으로 논평에 특정 인용 주석에 대한 평가적 입장이 명시된 경우에 한하여 인용에 의도적인 맥락성을 부여하였으며, 해당 인용과 논평을 함께 연결하는 방식으로 관계 데이터를 구성하였다. 또한 견해 내부에서 제시된 경학자의 해석과 인용 주석의 해석이 명확히 일치하거나 배치되는 경우에도 그 관계를 평가적 맥락으로 해석할 수 있다고 판단하여 인용 맥락을 부여하였다.

○ 『상서고훈』 예시

■ 긍정 인용 맥락(positive reference; PR)

- 요전2장5구: 黎民於變時雍
- 요전2장5구_훈고1인용: 梅云, 黎, 衆.
- 요전2장5구_훈고1견해1논평: 梅義, 似長也.

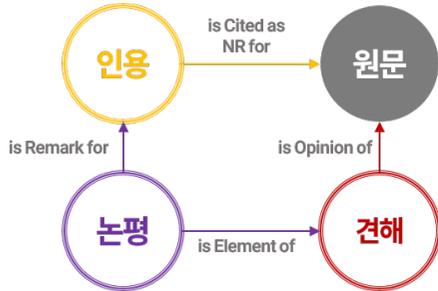
■ 부정 인용 맥락(negative reference; NR)

- 요전2장4구: 協和萬邦
- 요전2장4구_훈고1인용: 疏云, 於萬邦, 變言協和, 其實相通也.
- 요전2장4구_훈고1견해1논평: 疏義, 非也. 炎黃以來, 諸侯習用干戈, 互相侵奪, 堯以大德鎮撫諸侯, 使之協睦而納之軌度之中, 四朝一巡, 考其功績, 無有一邦或敢梗化, 茲所謂協和也, 豈可與平章通用?

■ 부분긍정 인용 맥락(partially positive reference; PPR)

- 요전1장1구: 曰若稽古
- 요전1장1구_훈고3인용: 蔡云, 曰若者, 發語辭. 史臣將敘堯事. 故先言考古之帝堯.
- 요전1장1구_훈고1견해1논평: 蔡說, 是也. 然讀至帝堯絕句, 猶夫未然.

[그림 3] 인용-원문 간 부정 인용(negative reference) 맥락을 표현하는 그래프 도상



견해-원문 관계는 경학자가 해당 원문에 직접적으로 대응한 부분이므로 자동형 서술 구조를 활용하여 “A is Opinion of B(A는 B의 견해이다)”의 관계어로 바로 연결하였다.

◎ 주석-주석 관계

주석 간 관계는 주석 정보에 해당하는 인용, 견해, 논평, 언급 클래스 간의 관계를 구조화한 것으로 다시 기본형, 내포형, 종속형의 세 유형으로 세분화할 수 있다.

기본형은 경학자에 의해 인용된 개별 주석들 사이의 관계성을 표현하기 위한 유형으로, 다양한 인용 간의 교차적 관계를 조직함으로써 지식의 계보성, 경학자의 주석 배치 의도 등을 심층적으로 파악할 수 있게 한다. 예컨대, 『사기』(史記)와 『사기색은』(史記索隱), 『춘추좌씨전』(春秋左氏傳)과 두예(杜預), 『국어』(國語)와 위소(韋昭) 등의 주석 관계나 주소(注疏) 관계 등 인용 주석 간의 명시적인 관계성이 있는 경우, 또 ‘박왈’(駁曰), ‘난왈’(難曰), ‘동’(同) 등 뚜렷한 관계 표지가 확인되는 경우를 기준으로 관계어를 설계하

였으며, 여러 사례를 유형화한 결과 보충 관계, 반박 관계, 동일 관계의 세 가지 관계로 구분하였다.⁵

보충 관계는 대상 주석을 해설하거나 풀이한 경우, 추가적인 정보나 배경 설명을 보충하는 경우를 모두 포괄하는 유형이다. 이 관계는 주로 주석 관계나 주소 관계에서 자주 확인되며, 후대의 학자가 선대 학자의 해석을 계승하는 맥락에서도 흔히 보인다. 반박 관계는 말 그대로 대상 주석에 이의를 제기하거나 반박하는 경우를 의미하며, 주로 ‘駁曰’, ‘難曰’ 등의 반박 표지와 함께 제시된다. 또한 『박오경이의』(駁五經異義)와 같이 특정 문헌의 해석에 전면적으로 반론을 제기하는 자료에서도 이러한 관계가 빈번하게 나타난다. 동일 관계는 ‘同’, ‘亦’ 등의 표지를 통해 확인할 수 있으며, 주로 동일 견해의 주석을 나란히 배치하여 논거를 강화하거나 해석의 정당성을 확보하려는 경학자의 의도를 반영하는 경우가 대부분이다.

○ 『상서고훈』 예시

■ 보충 관계(supplement)

- 요전11장1구_고증7인용: 鄭語史伯云, 姜, 伯夷之後也.
- 요전11장1구_고증8인용: 韋昭云, 伯夷, 堯秩宗, 炎帝之後, 四岳之族也.

■ 반박 관계(contradict)

- 요전4장6구_훈고2인용: 鄭云, 日中者, 日見之漏與不見者齊. 日長者, 日見之漏五十五刻, 於四時最長也. 夜中者, 日不見之漏與見者齊. 日短者, 日見之漏四十五刻 於四時最短也.

5 처음에는 주석 관계 및 주소 관계 등을 독립적인 관계 데이터로 반영하는 방안을 고려하였으나, 이들 관계가 일관되거나 단일한 성격으로 정의되지 않는다는 점에서 적용을 보류하였다. 주석-주석(기본형) 관계의 속성값으로 주석 관계 및 주소 관계와 같은 명시적 관계성을 반영하는 방향을 고려 중이다.

- 요전4장6구_훈고3인용: 王肅難云, 知日見之漏, 減晝漏五刻, 不意馬融爲傳, 已減之矣. 因馬融所減而又減之. 故日長爲五十五刻. 因以冬至反之, 取其夏至夜刻, 以爲冬至晝短, 此其所以誤耳.

■ 동일 관계(is Aligned with)

- 요전1장7구_훈고1인용: 梅云, 至于天地.
- 요전1장7구_훈고2인용: 蔡說, 同.

○ 『상서고주』 예시

■ 보충 관계(supplement)

- 요전5구_교감1인용: 漢藝文志, 合於堯之克攘.
- 요전5구_교감2인용: 注師古曰, 攘古讓字.

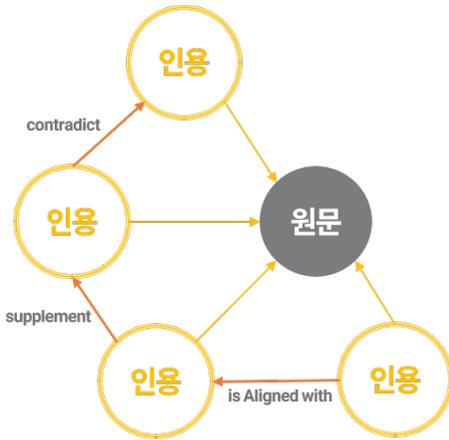
■ 반박 관계(contradict)

- 요전8구_훈고2인용: 許慎五經異義, [...] 明在九族中, 不得但施於同姓.
- 요전8구_훈고3인용: 鄭玄駁云, 異姓之服, 不過緦麻. 言不廢昏. 又昏禮請期云, 惟是三族之不虞, 恐其廢昏, 明非外族也.

■ 동일 관계(is Aligned with)

- 요전3구_훈고1인용: 馬融曰, 放勛堯名.
- 요전3구_훈고2인용: 趙歧孟子注, 同.

내포형은 주석자가 자신의 견해를 제시하는 과정에서 제3자의 주석을 다시 인용한 경우를 표현하기 위한 관계 유형으로, ‘인용 내 인용-인용’, ‘견해 내 인용-견해’ 관계가 이에 해당한다. 인용자가 제3자의 주석을 인용한다는 점에서 앞서 제시한 인용-원문 관계와 유사하지만, 이 유형은 주석과 원문 간의 층위가 아니라 주석과 주석 간의 관계를 구조화한다는 점에



[그림 4] 주석-주석(기본형) 관계를 표현하는 그래프 도상

서 구별된다. 해당 관계는 인용 목적에 따라 세 가지로 세분화하였으며, 이는 근거로 삼기 위해 인용한 경우, 반박하기 위해 인용한 경우, 참고를 위해 인용한 경우로 구성된다.

○ 『상서고훈』 예시

■ 근거 인용(is cited to support)

- 요전7장2구_훈고1인용: 王肅云, 改易者, 謹約蓋藏, 循行積聚. 詩云, 日爲改歲.
- 요전7장2구_훈고2인용: 詩云, 日爲改歲.

■ 반박 인용(is cited to contradict by)

- 요전2장1구_훈고1견해: 鋪案, 鄭意 [...]. 毛大可, 力主古訓, 非矣.
- 요전2장1구_훈고1견해2인용: 毛大可, 力主古訓, 非矣.

■ 참고 인용(is cited by)

- 요전2장_교감2인용: 史記索隱云, 今文, 平爲辯.
- 요전2장_교감3인용: 今文, 平爲辯.

○ 『상서고주』 예시

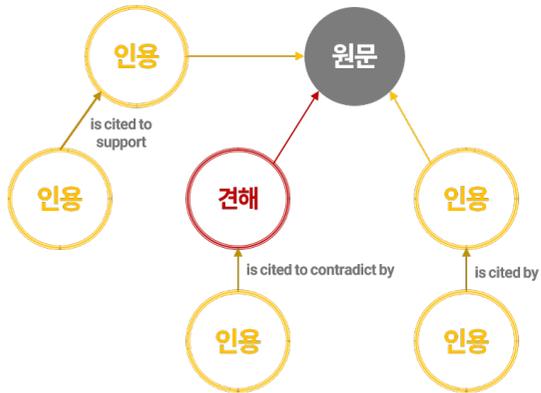
■ 근거(is cited to support)

- 요전43구_교감1견해: 綽按, 庸, 用也, 用可乎, 言不可用也. 史記, 頑凶不用, 卽馬義也.
- 요전43구_교감1견해1인용: 史記, 頑凶不用.

■ 참고 인용(is cited by)

- 요전11구_교감3인용: 師古曰, 今尙書作變, 此紀作蕃, 兩說並通.
- 요전11구_교감4인용: 今尙書作變.

[그림 5] 주석-주석(내포형) 관계를 표현하는 그래프 도상



마지막으로 종속형은 주석 요소를 구성하는 논평, 언급, 인용, 견해 클래스 간의 내부 영향 관계를 설계함으로써 경학자가 자신의 견해를 논리적으로 전개하거나 보완하기 위해 활용한 논증 방식을 구조화하기 위한 관계 유형이다. 이를 통해 하나의 견해 안에서 논거가 주체적으로 제시되었는지, 다른 주장과 어떠한 영향 관계를 맺고 있는지 등을 파악할 수 있다.

경학자가 주석을 인용한 맥락을 입체적으로 이해하기 위해 견해 내에 제시된 논평 부분을 ‘논평-견해’ 관계로 연결하고, 해당 논평이 대상으로 하는 인용을 다시 ‘논평-인용’ 관계로 연결함으로써 경학자의 주장과 논거

사이의 논리적 구조를 명확히 하였다. 또한 곳곳에 등장하는 언급을 해당 인용과 연결하여 견해 부분에서 확인할 수 없었던 경학자의 추가적인 해석 관점이나 태도를 파악할 수 있도록 하였다.

○ 『상서고훈』 예시⁶

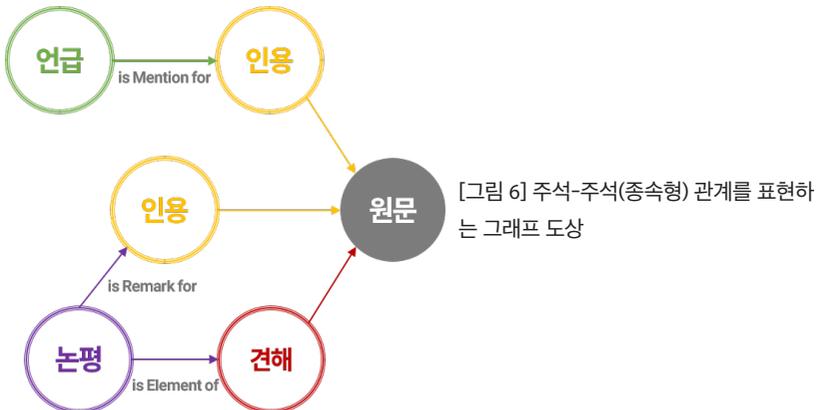
■ 언급-인용 관계(is Mention for)

- 요전7장2구_훈고1인용: 鄭駁云, 春氣博施. 故以廣大言之. 夏氣高明. 故以遠大言之.
- 요전3장2구_훈고5인용1언급: 評曰, 夏氣不可云高明.

○ 『상서고주』 예시

■ 언급-인용 관계(is Mention for)

- 요전20구_교감4인용: 說文, 作平黷.
- 요전20구_교감4인용언급: 後皆放此.



6 '논평-견해', '논평-인용'의 예시는 앞서 '◎ 주석-원문 관계' 부분에서 구체적으로 다루었다.

◎ 인용 내 관계

이 관계는 인용 부분을 구성하는 주요 정보 간의 관계를 연결하기 위한 것으로, 인용 구절의 출처가 문헌을 통해 제시되었는지, 인물(발화자)을 통해 제시되었는지를 고려하여 문헌과 인물 간의 관계를 세밀하게 구성하였다.

인용은 출처와 해당 구절로 구성되며, 이 두 요소는 모두 인용과 연결될 수 있다. 그러나 그래프의 연결성을 고려하여 먼저 구절-인용을 연결한 다음 구절을 매개로 하여 문헌-구절, 인물-구절의 관계를 추가로 설정하였다. 출처가 문헌으로 제시된 경우에는 해당 문헌의 저자를 다시 연결하였고, 인물로 제시된 경우 그 인물의 견해가 수록된 문헌과의 관계를 별도로 연결하였다.

구절-인용 간의 연결은 “A is Text of B(A는 B의 구절이다)”의 관계어로 연결하였다. 문헌-구절, 인물-구절 간의 연결은 모두 “A mention B(A가 B를 말하다)”의 관계어로 설정하였는데, 이는 ‘曰’, ‘云’ 등으로 출처와 구절을 연결하는 원문의 일반적인 인용 형태를 반영한 것이다.

출처가 문헌으로 제시된 경우 저자와 문헌 간의 관계는 “A is Writer of B(A는 B의 저자이다)”의 관계어로, 인물로 제시된 경우에는 수록 문헌과 발화자 간의 관계를 “A contain B(A는 B를 포함하다)”의 관계어로 각각 연결하였다.

문헌 제시 유형 중 인명과 서명이 함께 제시된 경우 해당 인물이 문헌의 저자이면 문헌을 기준으로, 발화자면 인물을 기준으로 관계를 연결하였다.



[그림 7] 구절-문헌, 구절-인물 연결 예시

추가로 XML 요소 내의 <Part>를 문헌-구절 관계의 속성값으로 기술하여 특정 구절이 문헌 내에서 어느 편목에 속하는지를 명시하였다. 이는 해석의 출처를 세밀하게 분석하려는 의도에서 비롯된 것으로, 편목 단위의 출처 분석을 염두에 두고 설계되었다.

◎ 주식-해석 관계

해석 정보 가운데 ‘재인용’의 경우 재인용 사례로 추정되는 인용 출처와 직접 연결하여 기존에 제시된 출처가 2차 자료를 통한 인용일 가능성을 반영하였다. 금고문은 인용, 견해, 논평, 언급 등 다양한 주식 단위에서 모두 나타나므로, 해당 클래스들과 각각 직접 연결하여 금고문 관련 담론을 통합적으로 추적할 수 있도록 설계하였다.

3. XML 파싱을 통한 그래프 데이터 구축과 그 논리

본 장에서는 온톨로지 디자인에 기반한 트리플 데이터를 편찬하기 위한 작업으로, 기구축 XML 데이터를 파싱(Parsing)하는 과정을 소개하고, 추출한 데이터를 시맨틱 데이터로 전환한 다음 그래프 데이터베이스 형태로 표현하는 방식을 다룬다. 특히 본 연구에서는 기존의 규칙 기반 파이썬 코딩 방식과 함께 최근 주목받고 있는 LLM을 활용한 파싱 방식을 병행하여 적용하였다. 파싱 과정을 통해 추출된 트리플 데이터는 그래프 데이터베이스 환경에서 관계 기반의 질의(Query)와 시각화가 가능하도록 정제되었으며, 이를 바탕으로 경학 자료에 내재한 복합적인 지식 체계를 연결망 형태로 재현하고자 하였다.⁷

7 현재 중국에서도 시맨틱 데이터 처리 기술을 활용하여 고전 자료를 지식 네트워크로 구조화하려는 시도가 활발히 진행 중이며, 지식의 추출·융합·계산·시각화를 통합적으로 다루는 그래프 데이터 기반의 지식 그래프 모델과 시맨틱 웹(Semantic Web) 기반의 링

3.1. 파이썬과 LLM(ChatGPT)을 활용한 파싱

온톨로지 기반의 트리플 데이터를 구축하기 위해 기구축 XML 문서를 대상으로 파싱 작업을 수행하였다. 데이터 파싱에는 파이썬과 LLM(ChatGPT) 두 가지 디지털 도구를 활용하였는데, 『상서고훈』은 파이썬, 『상서고주』는 LLM(ChatGPT)을 통해 각각 처리하였다.

먼저, 파이썬을 활용한 파싱을 과정을 소개한다.⁸ 각 클래스의 노드 사이트를 생성하기 위해서는 해당 클래스를 구성하는 정보에 대한 데이터 추출이 필요한데, 가장 많은 연결점과 속성을 지닌 Citation(인용) 노드 추출 과정을 전반적인 흐름을 중심으로 서술하겠다.

개별 노드 데이터를 본격적으로 추출하기 위한 예비 작업으로 깃허브(GitHub)에 탑재된 XML 파일을 불러온 후 content(내용) 값의 추출을 위한 처리를 진행하였다. XML상의 <Text> 태그 내부에 포함된 모든 텍스트 값을 빠짐없이 수집하기 위해 여러 하위 태그를 재귀적으로 탐색하고, 태그 종료 후에 등장하는 잔여 텍스트까지 함께 수집할 수 있도록 코드를 작성하였다. 이를 통해 <Text> 태그 전반에 걸쳐 존재하는 모든 텍스트 값을 계층 구조와 무관하게 평평한 리스트 형태로 추출할 수 있도록 처리하였다. 이상의 예비 작업을 수행한 후 id(식별자), name(대표명), area(영역), form(형태), content 값을 차례로 추출하여 일차적으로 사이트에 기술하였고, format(형식)과 context(맥락) 값을 자료를 독해하는 과정을 거쳐 차례로 입력하였다.

그 결과 추출된 데이터는 [그림 8]과 같다.

한편 『상서고주』는 ChatGPT4o를 활용하여 데이터를 추출하였다. LLM을 활용한 목적은 연구자가 다루기 힘든 방대한 데이터 속에서 프롬프팅을 통해 원하는 데이터를 목적에 맞게 추출 및 가공하는 것이다. 여기에서 관

크드 데이터(Linked Data) 모델이 함께 논의되고 있다. 陳濤 외(2019), pp. 34-44.

8 『상서고훈』 XML 문서와 전체 파싱 코드는 깃허브 계정 <COD-Project Repository> (URL: <https://github.com/KU-HJHM/COD-Project>)에서 확인할 수 있다.

기 위해 명령어 중 3번에서 상위 태그값을 의미하는 'Parent tag'를 출력하도록 하였다.¹⁰ 이후 직접 인용문인 <Text>를 추출하기 위해 ① <Text> 태그값이 무엇인지 명확히 정의하고 ② <Text> 내에 다른 태그가 있어서 직접 인용문이 온전히 추출되지 않는 현상을 방지하기 위해 다른 태그는 무시하고 추출하도록 명시하였다. 마지막으로 Citation 내에 직접 인용문이 아니어서 <Text>로 태깅되지 않은 간접 인용문도 함께 추출하기 위해 <Citation> 내에 있는 값을 전부 추출하도록 명령하였다.

그 결과 추출된 데이터는 [그림 9]와 같다.¹¹

다만 해당 데이터는 한 번의 명령에 의해 추출된 것이 아니다. 데이터 추출 과정에서 오류가 발생하여 수정사항을 다시 한 번 명령하였고 그 결과 위와 같은 온전한 데이터가 나올 수 있었다. ChatGPT를 사용하면서 주로 발생하였던 오류는 요청하지 않은 명령을 임의로 수행하는 경우, 자료를 오름차순으로 나열할 때 순서를 틀린 경우, 데이터를 누락시킨 경우, 데이터를 Excel 파일로 추출할 수 없다고 답한 경우 등이 있다.¹² 오류의 발생 원인은 아직 명확하지 않다. 다만 오류와 시행착오를 최대한 줄이고 원하는 데이터를 추출해내는 방법은 논리적 연산 과정을 최대한 상세하게 명시한 프롬프트를 제시하는 것이라 하겠다.

소를 기반으로 구분한 것이지 내용이 반드시 태그의 기준으로 명확히 구분되지는 않는다는 점을 밝혀 둔다.

- 10 XML은 상하 위계를 비롯한 각종 요소를 사용자의 필요에 따라 마크업할 수 있는 확장 가능한 언어로 특정 태그의 상위 요소를 'Parent', 하위 요소를 'Child'라 한다. 예시로 든 프롬프트에서는 XML 작업 시 통용되는 용어를 사용하여 LLM에게 사용자의 의도를 정확히 전달하고자 하였다.
- 11 명령어와 데이터 추출 과정은 해당 링크(URL: <https://chatgpt.com/share/67f5d1cf-a870-8010-87d4-bd1d335a6511>)에서 확인할 수 있다.
- 12 ChatGPT4o를 사용하여 데이터를 추출하는 과정에서 오류가 발생한 경우는 다음의 링크에서 확인할 수 있다.
- Excel 파일 추출 불가: <https://chatgpt.com/share/67c14275-a0b0-8010-8c9f-139e67545d02>
 - 데이터 누락: <https://chatgpt.com/share/67f50211-6dc4-8010-929d-65bb5e189f42>
 - 순서 틀림: <https://chatgpt.com/share/67f50758-7924-8010-803d-0df68f09a936>

상서고훈과 상서고주 노드 데이터 ☆ 📄 🔄

파일 수정 보기 삽입 서식 데이터 도구 확장 프로그램 도움말

🔍 메뉴 🏠 ↶ ↷ 🗨️ 🔊 100% W % 0.00 123 기본값... - 10 + B I U A 🗑️ 🌐 📄

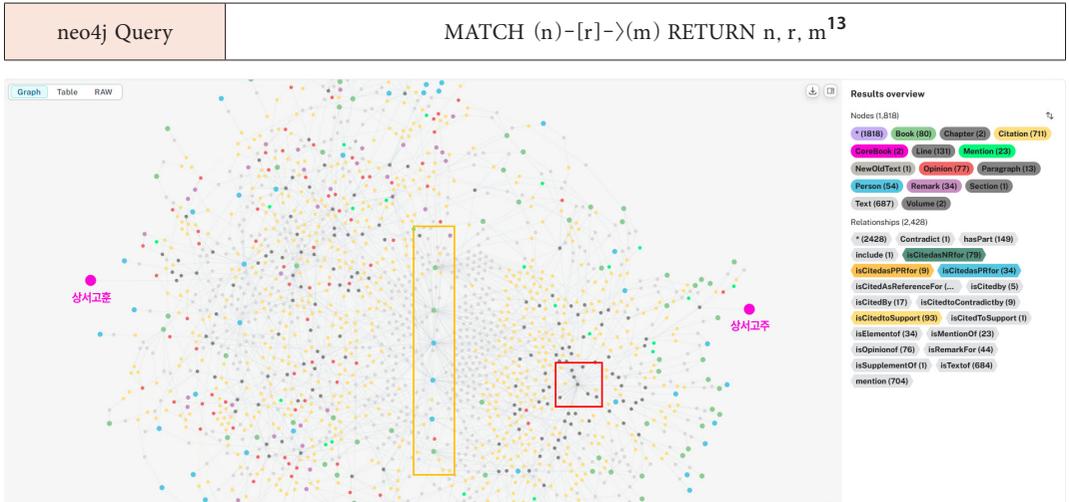
T466 | ㅈ

	C	D	E	F	G
434	상서고주				
435	id	name	class	area	content
436	식별자	대표명		영역	내용
437	S1_GC11	우하서_종고1인용	Citation	종고	鄭玄書贊虞夏書二十篇
438	S1_GC12	우하서_종고2인용	Citation	종고	孔穎達曰馬融鄭玄王肅別錄題皆曰虞夏書
439	S1_GC13	우하서_종고3인용	Citation	종고	又曰虞舜與禹相授事相嗣穿故尚書舜禹之書詳
440	S1_GO1C11	우하서_종고1견해1인용	Citation	종고	許慎說文引唐書曰五品不遜
441	S1_GO1C12	우하서_종고1견해2인용	Citation	종고	三國志陸抗疏稱諸葛回唐書依戒並以舜禹爲唐
442	S1_GO1C13	우하서_종고1견해3인용	Citation	종고	大傳雖曰虞夏傳而虞夏傳中亦別唐傳虞夏傳
443	S1_GO1C14	우하서_종고1견해4인용	Citation	종고	說文雖五品不遜稱唐書而其餘引舜禹者二十二處
444	S1_GO1C15	우하서_종고1견해5인용	Citation	종고	左傳引禹禘皋陶諱者五處而皆曰夏書是舜禹禹事
445	S1_GO1C16	우하서_종고1견해6인용	Citation	종고	王充曰唐虞夏殷禹者土地之名舜以唐侯嗣位舜從
446	S1_GO2C11	우하서_종고2견해1인용	Citation	종고	蔡傳總題于上
447	C01_GC11	요전_종고1인용	Citation	종고	劉熙曰九丘三王以前上古義皇時書今皆亡唯堯典
448	C01_CC11	요전_교감1인용	Citation	교감	大學引此舜作帝
449	C01L001_GC11	요전1구_종고1인용	Citation	종고	鄭玄曰稽古也古天也言能順天而行之與之同功
450	C01L001_GC12	요전1구_종고2인용	Citation	종고	三國志魏志帝幸太學問諸儒曰日稽古同天言舜同於
451	C01L001_GC13	요전1구_종고3인용	Citation	종고	文選王延壽賦考稽古注張載曰若順也稽考也若
452	C01L001_GC14	요전1구_종고4인용	Citation	종고	想譯新論最近君能說堯典二字十餘萬言但說曰若
453	C01L001_CC11	요전1구_교감1인용	Citation	교감	王延壽賦注李善引此作曰作考稽古
454	C01L001_CC12	요전1구_교감2인용	Citation	교감	鬼谷子考稽稽古聖人亦作考
455	C01L001_CC13	요전1구_교감3인용	Citation	교감	唐書落碑作考若古古亂古稽字
456	C01L002_GC11	요전2구_종고1인용	Citation	종고	馬融曰堯總也虞書傳聖曰堯
457	C01L002_GC12	요전2구_종고2인용	Citation	종고	孔穎達曰鄭于下傳云虞氏舜名則堯舜相配故舜禹
458	C01L002_GC13	요전2구_종고3인용	Citation	종고	劉熙曰魯古堯堯然故謂之堯也
459	C01L003_GC11	요전3구_종고1인용	Citation	종고	馬融曰放勳堯名
460	C01L003_GC12	요전3구_종고2인용	Citation	종고	禮疏孟子注同
461	C01L003_CC11	요전3구_교감1인용	Citation	교감	說文勳作勳
462	C01L003_CC12	요전3구_교감2인용	Citation	교감	尚書考靈鑿同
463	C01L003_CC13	요전3구_교감3인용	Citation	교감	五帝本紀帝堯者放勳
464	C01L003_CC14	요전3구_교감4인용	Citation	교감	注疏守節曰放上代之功勳
465	C01L004_GC11	요전4구_종고1인용	Citation	종고	馬融曰虞儀表備請之欵照臨四方講之明理維天地

[그림 9] 『상서고주』 Node_Citation 시트 일부

3.2. 시맨틱 데이터와 그래프 데이터베이스

데이터 파싱 작업을 통해 노드와 엣지를 구성하는 데이터를 편찬하고 나면 Neo4j 등의 그래프 데이터베이스를 활용하여 지식 그래프 형태로 출력하고, 다양한 질의를 해볼 수 있다. 경학 자료의 경우 자료에 내재한 지식 간의 상호작용과 구조적 입체성을 시각적으로 드러낼 수 있다는 점에서, 그래프 기반 데이터 표현은 테이블 형식의 표나 텍스트 기반의 독해 방식보다 훨씬 풍부한 분석 가능성을 제공한다. 또한 연구 질문을 반영한 사이퍼 쿼리의 생성을 통해 특정 학자 또는 문헌의 파급 범위, 해석의 계보, 논점의



[그림 10] 『상서고훈』과 『상서고주』 전체 데이터 출력 화면

중심성 등을 시각적으로 재현할 수 있다.¹⁴

[그림 10]은 『상서고훈』과 『상서고주』의 총 14개 클래스(원전 정보의 ‘절’ 제외) 1,817개의 노드 데이터와 19개 관계 유형 2,428개의 엣지 데이터를 입력하고 전체 데이터를 출력한 화면이다. 두 자료가 연계되는 부분은 인용의 출처에 해당하는 문헌(Book), 인물(Person) 클래스이며, 특히 노란색 사각형으로 표시한 노드들을 중심으로 높은 연결 가중치를 보인다. 연결점이 많은 상위 3개 노드는 정현(鄭玄)(77회), 『사기』(57회), 『설문해자』(說文解字)(48회)이며, 해당 노드를 중심으로 문헌 및 학자의 영향 범위를 확장하여 살펴볼 수 있다.

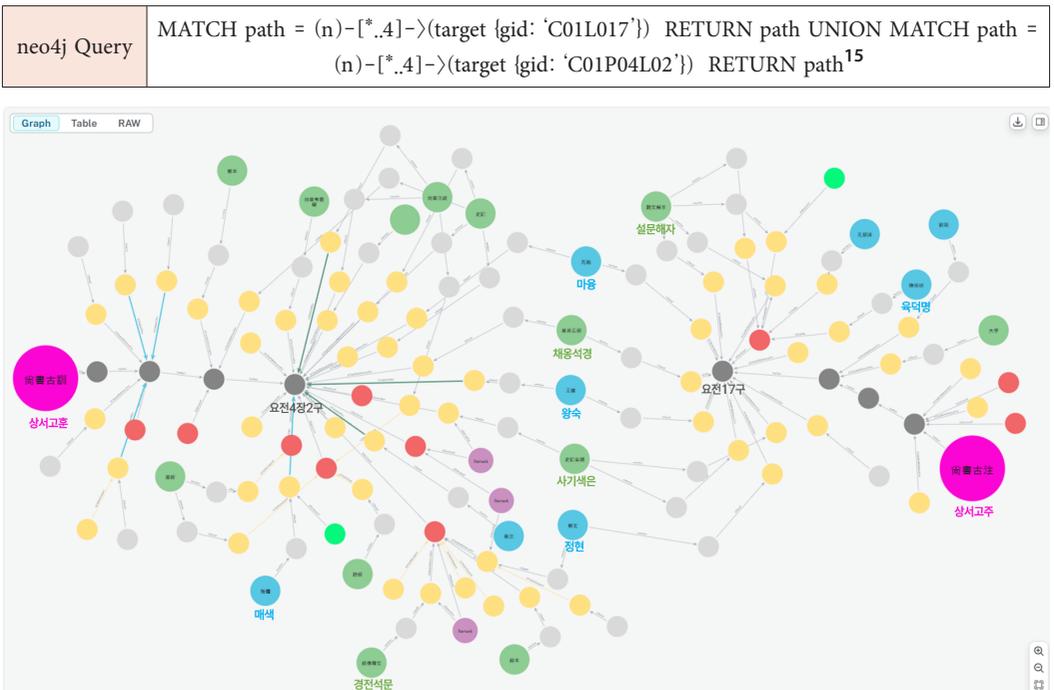
또 『상서고주』의 경우 ‘편’ 밑에 바로 ‘구’를 배치하였으므로, 「요전」을 중심으로 한 하위 구들 간의 연결점이 『상서고훈』에 비해 상대적으로 밀집되어 있으며, 이는 붉은색 사각형으로 표시한 영역을 통해 확인할 수 있다.

13 그래프 내의 모든 방향성이 있는 관계와 관련 노드들을 전부 불러오는 사이퍼 쿼리.

14 본 연구에서는 <Neo4j Aura>(URL: <https://neo4j.com/product/auradb/>)를 활용하였다.

그리고 『상서고훈』이 『상서고주』에 비해 저자의 견해를 적극적으로 제시하였다는 점이 붉은색 Opinion 노트, 자두색 Remark 노트의 분포를 통해 확인된다.

특정 구절에 대한 두 학자의 문헌고증 방식을 살펴보는 것도 가능하다. [그림 11]은 ‘택우이’(宅隅夷) 구에 대한 『상서고훈』과 『상서고주』의 주석 관계망을 시각화한 화면이다.



[그림 11] ‘택우이’(宅隅夷)에 대한 『상서고훈』과 『상서고주』의 주석 관계망

‘택우이’ 구는 『상서고훈』에서는 「요전」 4장 2구, 『상서고주』에서는 「요전」 17구에 해당한다. 다산의 자료 활용이 상대적으로 더 다양하게 나타난

15 ‘宅隅夷’를 content로 가지는 『상서고훈』의 「요전」 4장 1구(C01P04L02) 노트와 『상서고주』의 「요전」 17구(C01L017) 노트 각각으로 향하는 경로를 최대 4단계 이내에서 추적한 후 이를 결합하는 사이퍼 쿼리.

다는 점을 직관적으로 파악할 수 있으며, 동시에 두 자료 모두 마융(馬融), 왕숙(王肅), 정현, 그리고 『채옹석경』(蔡邕石經)과 『사기색은』을 공통적으로 인용하였음도 확인할 수 있다. 다만 다산은 왕숙의 해석을 부정하는 맥락에서 인용하였고(초록색 화살표), 석천은 별다른 평가적 입장 없이 단순 인용하였다는 점에서 차이를 보인다. 또한 동일 구절을 인용한 사례로, 『상서고훈』에는 『경전석문』, 『상서고주』에는 육덕명(陸德明)이 출처로 제시되는 등 표기 방식에 차이를 보이는 경우도 일부 확인하였다. 또 다산은 훈고 영역에서 매색(梅賾)의 주석을 긍정 인용하여 지지하였으나(하늘색 화살표), 『상서고주』에서는 해당 주장이 따로 인용되지 않았다는 점도 살펴볼 지점이다.

본 구는 이설이 다수 존재했던 만큼 두 자료 모두에서 다양한 해석 활동이 전개되었다. 『상서고훈』의 경우 견해를 제시한 부분이 다섯 곳이나 되며, 인용 구절에 대한 ‘논평’도 세 차례 확인된다. 『상서고주』 역시 신작 본인의 견해가 제시된 부분이 있으며, 견해 내의 지식 구성에서 『설문해자』의 활용이 특히 두드러진다.

이상과 같이 구를 중심으로 두 자료가 연계하는 양상을 확인함으로써 자료 간 공통되는 지점과 상이한 해석적 접근이 보이는 지점을 제시할 수 있으며, 이러한 단서들을 해석하고 확장해 살펴봄으로써 경학 자료 간의 연계 비교 연구 가능성을 구체적으로 모색할 수 있다.

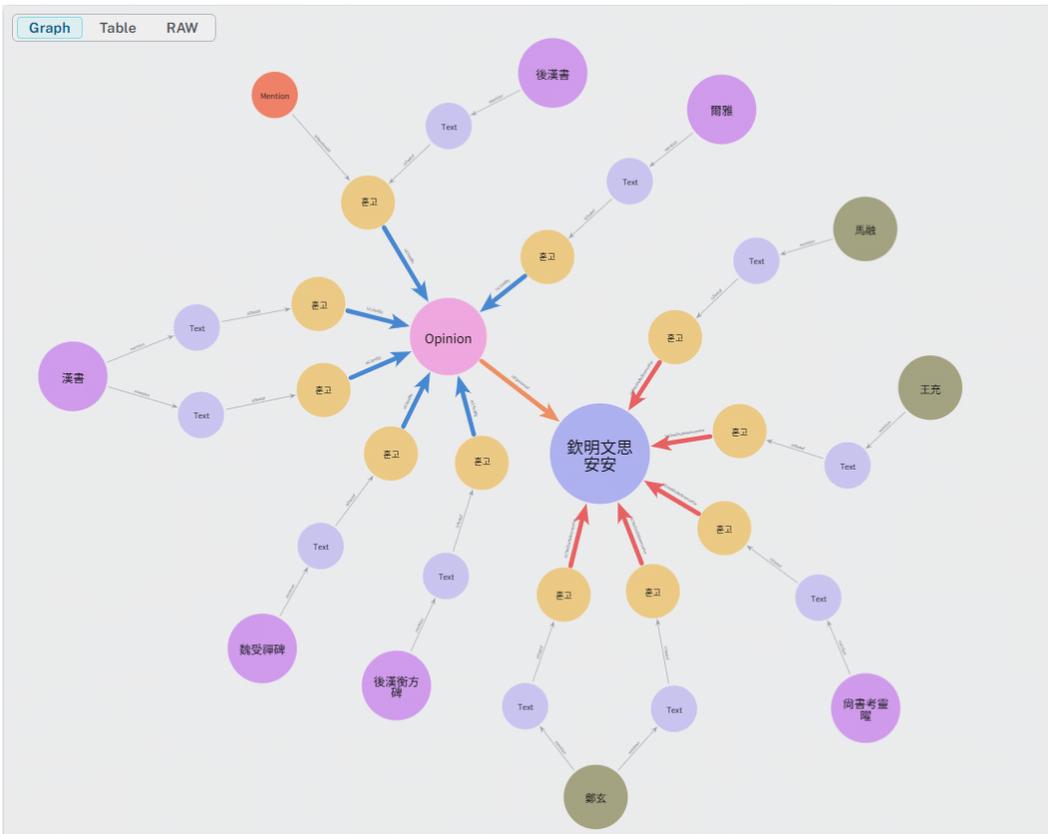
다음으로, 개별 자료를 대상으로 한 그래프 데이터베이스 구현 예시를 소개하겠다. 『상서고훈』의 경우 구마다 다산이 자료를 활용한 양상을 세밀히 살펴보는 것이 중요하다. [그림 12]는 요전 1장 1구 ‘왈약계고’(日若稽古)에 대한 『상서고훈』의 주석 관계망을 시각화한 것이다.

‘왈약계고’에 대한 기존의 해석은 후대에 와서 많은 비판을 받았는데, 다산 역시 기존의 대표 고주(古注)인 매색(梅賾)의 주장 등을 부정하며(초록색 화살표) 본인의 견해를 제시하였다. 추가로 ‘고정’이라는 분류표지를 활용하여 기존 『사기』, 『한서』(漢書) 등의 주석도 비판하였으며, 매색의 다른 주장을 다시 인용하기도 하였다. 또 정설 가운데 비교적 이른 시기에 제시된

하기도 하였다.

이처럼 다산은 하나의 구에 대해 매우 다양한 인용 서적을 활용하였고, 논평, 언급과 같은 주석 방식 및 고정 등의 분류표지를 사용하여 고증을 수행하였다. 이러한 주석 활동의 복잡성을 고려할 때 그래프 데이터베이스를 활용한 구현 방식은 다산의 다층적인 인식 구조와 주석의 논리적 연결망을 드러내는 데 유효한 방법이 될 수 있다.

neo4j Query	MATCH path = (n)-[rels*..6]->(t:Line {gid: 'C01L004'}) WHERE ALL(r IN rels WHERE type(r) <> 'hasPart') RETURN path ¹⁷
-------------	--



[그림 13] 「요전」 4구에 대한 「상서고주」의 주석 관계망

[그림 13]은 요전 4구 ‘흠명문사안안’(欽明文思安安)에 대한 『상서고주』의 주석 관계망을 지식 그래프 형태로 시각화한 것이다. 석천은 이 구절에 대해 훈고 영역에서 총 5개의 주석을 인용하고 견해(Opinion)를 제시하였다. 훈고는 정현, 마용, 왕충(王充), 『상서고령요』(尙書考靈曜)를 인용하였다. 견해는 석천 본인의 직접적 발화가 아닌 후한형방비(後漢衡方碑), 위수선비(魏受禪碑), 『한서』, 『후한서』, 『이아』(爾雅)에서 인용문을 가져와 나열한 방식으로 구성되어 있다. 이상의 지식 그래프를 통해 석천은 훈고 영역에서 5개의 주석을 가져왔고, 다시 견해에 내포된 인용문을 통해 추가 설명을 가한 구조임을 알 수 있다.

예시로 든 4구를 비롯하여 『상서고주』 「요전」에서는 인용한 주석에 대한 석천의 비판적 견해 혹은 논평 등이 보이지 않는다. 이는 석천이 경문을 충실히 설명하면서 본인의 견해를 대변하는 주석을 정제하여 수록하는 방향으로 저술하였기 때문이다. 그럼에도 불구하고 작안(續按)이라는 표지를 두어 본인의 견해가 개입되었음을 명확히 밝히고 있는데, 이는 추후 『상서고주』의 전체 데이터를 입력하고 그중 견해(Opinion)들을 추출하여 모아본다면 인용된 주석과 작안의 층위를 보다 뚜렷하게 파악할 수 있을 것으로 보인다.

이상의 두 사례는 특정 구절을 중심으로 주석 요소 간의 상호 관계망이 어떻게 구성되어 있는지를 시각적으로 구현한 것이다. 구절을 중심으로 재현한 관계망 외에도 특정 문헌이나 인물, 쟁점 등을 중심으로 관계망을 재구성함으로써 주석자의 자료 활용 방식이나 해석의 계보, 문헌 간 상호 참조 구조, 인용의 분포와 밀도 등을 다각적으로 분석할 수 있는 가능성을 제시한다.

17 ‘C01L004’인 Line 노드를 도착점으로 설정하고, 이 노드로 향하는 방향의 관계를 따라 최대 6단계 깊이에서 연결된 상위 노드들과의 경로를 추적하되, hasPart 관계를 포함하지 않는 경로만을 선별하여 시각화하는 사이퍼 쿼리.

4. 맺음말: 데이터를 매개한 질적 연구와 양적 연구의 균형

자료를 데이터로 표현하는 작업은 단순히 구조화된 정보를 편찬하는 기술적 과정에 그치지 않는다. 원전을 독해하면서 얻게 된 지식을 글이 아닌 데이터로 기술하는 것은 머릿속에서 사고를 정리한 후 문장으로 서술하는 기존의 방식과는 다른 차원의 접근을 요구한다. 해석의 단편들을 데이터로 기술하고 이를 지속적으로 확장하여 활용할 수 있다는 점에서 기억의 한계를 보완하고 개인의 사고를 넘어서는 분석의 가능성을 제공한다.

데이터를 설계하고 가공하며 검수하는 일련의 과정 자체가 정성적 독해의 연장이라고 할 수 있다. 복잡한 지식 간의 유기적 관계를 파악하고 이를 연결함으로써 텍스트에 내재한 다양한 의미의 층위를 반영하려는 시도는 질적-양적 연구를 아우르는 작업이었다.

특히 정확 자료를 중심으로 경문과 주석, 그리고 주석 간의 관계를 파악하는 과정은 텍스트에 대한 세밀한 독해를 요구했다. 그러나 동시에 자료의 지식을 데이터로 표현하는 과정에서 객관적 기준과 해석적 판단 사이의 조율이 필요했다. 텍스트 내에 명시적 표지가 없는 경우 연구자가 임의로 관계를 추정하여 데이터에 반영하게 되면 주관적 해석이 개입될 위험이 있었다. 결국 뚜렷한 표지가 없는 경우에는 해석적 맥락을 반영하지 않는다는 원칙을 세웠다. 이는 데이터 편찬의 일관성을 유지하기 위한 것이었으나, 결과적으로 데이터의 입체성을 오히려 제한하는 결과를 초래하지 않았는지에 대한 의문이 남아 있다.

예컨대 『상서고주』의 경우 석천은 경문 해석에 있어 자신의 견해를 드러내기보다 다양한 주석을 병렬적으로 나열하는 방식을 택하고 있어, 개별 주석이 지닌 해석적 함의를 파악하는 데 어려움이 있었다. 「요진」 1구 ‘월약계고’ 구절을 둘러싸고 정현, 황계, 가규(賈逵), 마옹, 왕숙, 그리고 『문선』(文選)에 이르기까지 여러 주석이 등장하지만, 뚜렷한 표지가 없이 병렬적

으로 나열되어 있으므로 신작이 『문선』의 인용을 통해 앞선 해석들을 절충하려 한 것인지, 단순 참고를 위해 인용한 것인지에 대한 판단은 해석자의 추론에 의존할 수밖에 없었다. 그 의도를 추론하여 데이터에 반영하는 방안을 고민해 보았으나, 객관성을 담보해야 할 데이터에 연구자의 주관이 개입될 위험이 있어 반영하지 않았다. 마찬가지로 ‘이친구족’(以親九族)에 대한 난외주 형태의 『백호통』(白虎通) 인용이나 ‘만방’(萬邦)에 대한 왕충의 해석 역시 인용의 성격과 맥락이 명시되어 있지 않아 데이터 표현 시 신중한 판단이 요구되었다. 『상서고훈』의 경우에도 비슷한 고민의 지점이 존재한다. 다산의 경우 비교적 상세하게 본인의 견해를 제시해 놓았지만, 견해의 훈고와 제시된 인용의 훈고가 배치된다는 것이 문면에 드러나더라도 다산의 논평을 통한 평가적 입장이 확인되지 않는다면 의도적 인용 맥락을 데이터에 적극적으로 반영해도 되는지에 대한 신중한 접근이 요구되었다. 즉, 명시적 견해가 나타나지 않는다면 이를 파악하기 위해 연구자는 해석적 추론을 하게 되는데, 그 추론을 어느 정도까지 데이터에 반영할 수 있는지는 여전히 고민의 지점으로 남아 있다.

이러한 고민은 데이터 편찬을 통한 지식의 표현이 질적 연구와 양적 연구의 경계를 넘나드는 복합적 작업임을 보여준다. 연구자의 주관을 최소화하면서도 해석적 통찰을 포기하지 않기 위해 끊임없이 ‘표현 가능한 것’과 ‘표현 불가능한 것’ 사이의 경계를 탐색하고 조율해야 했다.

이러한 난점을 마주한 가운데, 인용 구절 간의 상호작용과 구조적 입체성을 시맨틱 기반의 논리적 구조로 제시함으로써 의미 기반 정보 연결의 가능성과 시맨틱 데이터의 잠재적 활용성을 실험적으로 검토하였다는 점에서 이번 연구의 의의를 찾을 수 있다. 또한 『상서고훈』과 『상서고주』 두 자료의 연계 지점을 탐색함으로써 비교 연구로의 확장 가능성을 제시하였다. 그래프 데이터베이스 분석을 위한 사이퍼 쿼리 생성 과정은 연구자의 해석적 관점과 식견이 반영되는 질적 연구의 영역으로 데이터 해석 및 활용 가능성에 결정적인 영향을 미치며, 앞으로의 연구 과정에서 주력해야 할

핵심적인 작업이다.

향후 시맨틱 데이터의 편찬이 일정 수준 이상 이루어지면 개별 자료를 대상으로 한 분석 연구를 수행하고자 한다. 원래는 자료 간의 비교 분석을 중점적으로 수행한 연구를 세 번째 논문으로 연이어 제출할 계획이었으나, 각 자료를 대상으로 생성하려는 연구 질문이 상이하다는 점을 확인하게 되었다. 이에 따라 우선 개별 자료에 대한 심층 연구를 진행하고, 각각의 연구 결과를 정리하여 두 편의 논문을 제출할 예정이다. 이후 두 자료를 통합하여 연계 분석을 시도함으로써 각 주석서의 해석 경향성과 문헌고증 방법론 및 자료 간의 영향 관계 등을 비교 연구하고, 이 내용을 『상서고훈』과 『상서고주』 연계 데이터 프로세싱의 세 번째 논문으로 제출할 계획이다.

참고문헌

- 김현·김바로·임영상(2016), 『디지털 인문학 입문』, HUEBOOKs.
 벤자민 엘먼 지음, 양희웅 옮김(2004), 『성리학에서 고증학으로』, 예문서원.
 劉起鈺 지, 이은호 역(2016), 『상서학사: 상서에 관한 2천여 년의 해석사』, 예문서원.
 류인태 외(2023), 『디지털로 읽고 데이터로 쓰다』, 박문사.
 성균관대학교 대동문화연구원(1998), 『朝鮮後期 經學의 展開와 그 性格』, 성균관대학교 출판부.
 왕요남 지음, 신승운 외 옮김(2014), 『주석학개론』, 한국고전번역원.
 정약용·신작 지, 실시학사경학연구회 편역(2000), 『茶山과 石泉의 經學論爭』, 한길사.
 정약용 지음, 실시학사경학연구회 옮김(2020~2022), 『다산 정약용의 상서고훈 1-5』, 학자원.
 정양완·심경호(1999), 『江華學派의 文學과 思想 4: 石泉 申紳論』, 한국정신문화연구원.
 黃俊傑 지, 최영진·안유경 역(2009), 『동아시아 유교경전 해석학』, 문사철.
 혼다 시게유키 지, 서경요 역(2011), 『중국경학사』, 문사철.
- 김문식(2013), 「다산 정약용의 『尙書』 주석서 계열 고찰」, 『다산학』 23, 다산학술문화재단, pp. 85-132.
 김바로(2018), 「디지털 인문학의 관점에서 본 디지털 불교학」, 『불교학보』 86, 동국대학교 불교문화연구원, pp. 265-288.

- 김바로(2023), 「한국역사인물데이터베이스 설계 시론: 홍문록(弘文錄)을 예시로 하여」, 『한국학』 46, 한국학중앙연구원, pp. 159-195.
- 김보름(2021), 「다산 정약용의 『상서고훈』 저술과 개정 과정 연구: 필사본의 검토를 중심으로」, 『유학연구』 56, 유학연구소, pp. 107-136.
- 김성재(2018), 「『상서고훈』의 구성 체제와 안설(案說)에 대한 고찰」, 『한국학연구』 48, 인하대학교 한국학연구소, pp. 521-552.
- 김성재(2021), 「家藏本 『상서고훈』과 『상서지원록』의 서지적 고찰」, 『다산학』 39, 다산학술문화재단, pp. 167-210.
- 김인중(2012), 「茶山の 經學 그 大略과 意義」, 『한국한문학연구』 50, 한국한문학회, pp. 37-70.
- 김지선·류인태(2022), 「여향문화 연구와 데이터 모델링」, 『한국한문학연구』 85, 한국한문학회, pp. 79-118.
- 김현(2018), 「디지털 시대의 漢文學: 데이터로 소통하는 고전 인문 지식」, 『한문학논집』 49, 근역한문학회, pp. 9-42.
- 김현·안승준·류인태(2018), 「데이터 기반 인문학 연구 방법의 모색: 문중 고문서 아카이브와 디지털 인문학의 만남」, 『횡단인문학』 창간호, 숙명여자대학교 숙명인문학연구소, pp. 17-61.
- 박순(2017), 「고전문학 자료의 디지털 아카이브 편찬 연구: 누정기(樓亭記) 자료를 중심으로」, 연세대학교 대학원 박사학위논문.
- 박승일(2024), 「『언어적 전회』에서 ‘인공적 전회’로: 합성곱 신경망과 오토인코더를 중심으로」, 『커뮤니케이션 이론』 20, 한국언론학회, pp. 48-102.
- 변은미(2022), 「다산 『상서고훈』의 훈고학적 접근」, 『경학』 4, 한국경학학회, pp. 1-37.
- 변은미·이동학·유인태(2024), 「『尙書古訓』과 『尙書古注』 연계 시맨틱 데이터 프로세싱 1: XML 데이터 설계 및 편찬」, 『한문학논집』 69, 근역한문학회, pp. 287-338.
- 선미현(2023), 「石泉 申緯의 詩經學 研究」, 성균관대학교 박사학위논문.
- 심경호(1991), 「다산 상서론에 관하여」, 『한국한문학연구』 19, 한국한문학회, pp. 365-381.
- 심경호(2012), 「다산 정약용의 문헌해석방법과 필롤로지」, 『다산학』 21, 다산학술문화재단, pp. 7-63.
- 양승목·유인태(2022), 「야담의 데이터, 야담으로부터의 데이터: 한국 야담 데이터 모델의 구상」, 『한국문학연구』 68, 동국대학교 한국문학연구소, pp. 275-306.
- 유인태(2019), 「데이터로 읽는 17세기 체지사족의 일상: 『지암일기(1692-1699)』 데이터베이스 편찬 연구」, 한국학중앙연구원 박사학위논문.
- 유인태(2022), 「데이터로 고전을 읽는다는 것」, 『고전문학과 교육』 51, 한국고전문학교육학회, pp. 41-70.
- 이길환(2024), 「한국한문소설 시맨틱 데이터 모델링에 대한 연구: 한국한문소설 DB 구축을 위한 예비적 검토」, 『민족문학사연구』 85, 민족문학사연구소, pp. 83-121.
- 이길환·변은미·유인태(2024), 「조선시대 料試 자료와 시맨틱 데이터 모델링」, 『한국한

- 문학연구』 92, 한국한문학회, pp. 65-104.
- 이상욱(2010), 「정약용 『상서고훈』의 의례(義例)와 존고(存古)의 문제」, 『한국한문학연구』 45, 한국한문학회, pp. 41-77.
- 이영호·함영대(2018), 『디지털 경전주석학의 모색: 한국경학자료시스템을 중심으로』, 『대동문화연구』 101, 대동문화연구원, pp. 35-66.
- 임현규(2015), 「다산의 『서경』 인식」, 『동양고전연구』 59, 동양고전학회, pp. 297-324.
- 장문석·유인태(2021), 「디지털 인문학과 한국문학 연구(1): 작가 연구를 위한 시맨틱 데이터베이스 설계」, 『민족문학사연구』 75, 민족문학사학회, pp. 347-426.
- 장지훈(2025), 「『설담집』 「봉행록」의 시맨틱 데이터 모델 설계」, 『한국문학연구』 77, 한국문학연구소, pp. 49-102.
- 지영원·최진경(2024), 「한국 한시 데이터 아카이브 구축을 위한 개념적 데이터 모델링 시론」, 『민족문학사연구』 85, 민족문학사연구소, pp. 43-82.
- 陳濤·劉偉·單蓉蓉·朱慶華(2019), 「知識圖譜在數字人文中的應用研究」, 『中國圖書館學報』 45(244), 北京: 國家圖書館出版社, pp. 34-44.
- 진역령(2016), 「『古文尙書』에 대한 조선시대 학자들의 인식과 고변」, 고려대학교 박사학위논문.
- 한희연·지영원·호정·류인태(2022), 「한국 詩話 시맨틱 데이터 아카이브 구현 방안 모색: 『詩話叢林』 대상 개념적 데이터 모델링을 중심으로」, 『한문학논집』 63, 근역한문학회, pp. 105-146.

<COD-Project Repository> URL: <https://github.com/KU-HJHM/COD-Project>

<한국고전종합DB> URL: <https://db.itkc.or.kr/>

<한국경학자료시스템> URL: <http://koco.skku.edu/>

<한국문집총간 XML 파싱(1)> URL: https://youtu.be/0zrMo0vJKXg?si=IXEkZmKIplj4iF_d

<한국문집총간 XML 파싱(2)> URL: https://youtu.be/Ll__1BqzSpM?si=Nl2pmwbHUISWeCeL

원고 접수일: 2025년 4월 10일, 심사완료일: 2025년 5월 5일, 게재 확정일: 2025년 5월 8일

ABSTRACT

Semantic Data Processing of
Sangseogohun and
Sangseogoku, Part 2

Byeon, Eun mi*

Lee, Dong hak**

Ontology Design and Parsing

This paper presents the second study on the semantic data processing of *Sangseogohun* (尙書古訓) and *Sangseogoku* (尙書古注). The study explores the process of converting semi-structured XML data into ontology-based triple data. Based on the structural characteristics of Confucian classics, as well as key contextual and interpretative information, this study designs an ontology focused on semantic relational links. A methodological framework is proposed to structurally analyze the interpretive logic and inter-commentary interactions embedded in Confucian texts, by means of data parsing and knowledge graph implementation.

The conceptual structure encompasses textual, contextual and interpretative information, grounded in the shared design logic of XML schema and Ontology design. A total of 15 Classes are defined to capture the variations in data linkage between XML and Ontology. In particular,

* Ph.D. Candidate, Department of Korean Language and Literature, Korea University (First Author)

** Ph.D. Candidate, Department of Korean Language and Literature, Korea University (First Author)

relationships among commentaries are categorized into three types — basic, nested, and dependent — to systematically reflect inter-citation patterns and the internal logical structures of commentaries. In addition, contextual information is linked through relation attributes to articulate semantic connectivity and interpretive lineages within the dataset.

XML data are parsed using Python and LLMs to extract Nodes and Edges, thereby laying the foundation for building a graph database. Based on this, a partial knowledge graph is implemented to visualize the interpretive structure of Confucian texts, demonstrating the potential of semantic data modeling for knowledge integration and cross-textual comparison in Confucian studies.

Keywords Confucian Studies, Sangseo, *Sangseogohun*, *Sangseogoju*, Jeong Yakyong, Shin Jak, Digital Humanities, Semantic Data, Ontology, Knowledge Graph, XML Parsing, Triple Data