

2024년도 연구계획서

한국학중앙연구원

연구과제명	국 문	고문헌 데이터 설계를 위한 선행 모델 분석 및 교육 콘텐츠 구축				
	영 문	Analysis of Prior Models and Construction of Educational Content for the Design of Historical Document Data				
개설과목명	고문헌자료정보화연구					
연구책임자	성 명	김바로				
	소속(직명)	문화예술학부(조교수)				
	전공분야	인문정보학·인문지리학 (인문정보학)				
	생년월일	1982년 11월 04일				
	주 소	[03459] 서울특별시 은평구 은평로 160, 1505호(응암동, 경향렉스빌)				
	연 락 처	010-7399-2046				
	이 메 일	ddokbaro@gmail.com				
공 동 연 구 자						
성 명	소 속	직 위	주 전 공	연 락 처 (전 화)	연 락 처 (이메일)	비 고

고문헌 데이터 설계를 위한 선행 모델 분석 및 교육 콘텐츠 구축

I. 연구 계획

1 연구의 목적 및 배경

본 연구과제는 인문학 연구의 중요 원천인 고문헌 자료를 왜곡과 오류 없이, 나아가 디지털 매체의 장점을 극대화하면서 전산화할 수 있는 효율적인 데이터 모델을 탐색하는데 목적이 있다. 본 연구 과제에 참여하는 연구원(수강생)들은 <고문헌자료정보화연구> 강의에서 인문 고전 문헌 텍스트의 원문을 보존하면서 인문학 정보와 사유를 부여할 수 있는 기계가독형 데이터로 전환 가능한 XML(eXtensible Markup Language)¹⁾에 관한 전문 지식을 습득한다. 또 역사학·한문학·불교학 분야의 선행 고문헌 전산화 사례들을 리뷰하고, XML을 기반으로 동작하는 디지털 텍스트 기술(記述) 규약인 TEI(Text Encoding Initiative)도 살펴본다. 이를 바탕으로 한국학에서 중요 연구자료로 활용되는 고문헌 텍스트를 디지털 판본으로 전환하는 데 필요한 적절한 지식과 경험을 학습하려 한다. 아울러 그 성과를 한국학 연구자를 위한 디지털인문학 교육 자료로 제작하여 위키독스(WikiDocs)와 유튜브(Youtube)를 통해 고적 전산화에 관심있는 연구자 및 일반 대중과 공유할 예정이다.

1967년 하버드 대학 에드워드 와그너(Edward Wagner) 교수의 '문과방목(文科榜目) 프로젝트' 이래 주요 한국학 자료들을 디지털화하려는 노력은 50여 년 이어져 왔다. 고적 전산화 프로젝트가 국가사업으로 추진된 지도 이미 20여 년이 넘었으나 종이 매체에 기록된 텍스트를 온라인 매체로 단순 전사(轉寫)하는 데 그치는 경우가 많았다.²⁾ 현재까지의 고전 자료 정보화는 디지털 환경에 맞는 최적의 데이터 편찬을 추구하기보다는 서적으로 간행된 자료의 형식과 내용을 고스란히 온라인 상에 옮기는 데 치중해왔다.³⁾ 텍스트 전달의 매체가 달라졌는데도 종이 매체의 한계를 그대로 계승하여 디지털 매체가 갖는 장점을 충분히 구현하지 못한 것이다.⁴⁾ 단순히 아날로그 매체를 디지털 매체로 전환하는 디지털화(digitalization)를 넘어서, 고문헌에 담긴 지식의 총체를 디지털 플랫폼 상에서 재구축할 수 있는 디지털 전환(digital transformation)이 필요하다. 김현은 이를 '문헌

1) XML(eXtensible Markup Language): 문서의 구조적 형식과 내용 요소들을 컴퓨터가 식별할 수 있는 명시적 정보로 제공할 수 있는 전자문서 마크업 언어. 1998년 W3C(World Wide Web Consortium)가 최초의 권고안이 출판되었으며 2006년 제2판인 버전 1.1로 갱신되었다.

2) 이는 고적 전산화의 노력이 1998년 실업구제와 경기진작을 목적으로 하는 정보화 근로사업의 일환으로 시작된 것과도 관련이 있다. 사업이 IMF 이후 '고학력 실업자 구제'라는 사회복지 시책의 차원으로 시작되면서 자료 전산화의 양적 증대만을 추구했던 것이다. 이듬해인 1999년에는 국가적으로 보존 및 이용 가치가 높은 지식정보를 디지털화하는 '지식정보자원관리사업'이 시행되었다. 정보화 대상 자원의 40% 이상이 우리 역사 및 문화와 관련이 깊은 고전 문헌일 정도로, 고문헌 자료의 디지털화는 지식 정보 산업의 중요 부분으로 인정받았으나 고문헌을 어떻게 새로운 디지털 매체에 재현할 것인가에 대한 논의는 여전히 부재했다(김현, 「고문헌 자료 XML 전자문서 편찬 기술에 관한 연구」, 『고문서연구』 29, 2006, 186쪽).

3) 김현, 「데이터 기반 漢文古典 研究를 위한 試論」, 전통문화연구회 창립 30주년 기념 학술회의 발표문, 2018, 4쪽.

4) 김바로, 「역사기록의 전자문서 편찬방법 탐구: 역사요소를 중심으로」, 『열상고전연구』 51, 2016, 204-205쪽.

의 정보화'에서 '지식의 정보화'로 명명하기도 했다.⁵⁾

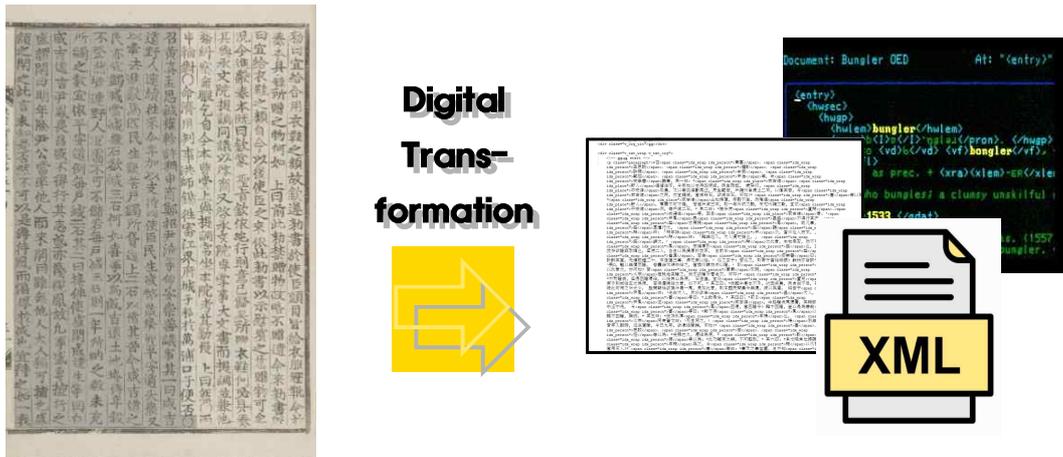


그림1) 고문헌 텍스트의 디지털 전환

마크업 언어(markup language)는 원문 보존을 충족하면서도 정확한 의미전달과 풍부한 해석의 가능성을 동시에 지닌 디지털 고문헌 판본 구축에 가장 핵심적인 정보 기술이다. 특히 XML(extensible markup language)은 고전 문헌과 같은 선형적이면서 비정형적 데이터를 기계가독형 데이터로 전환시키는 가장 적절한 방법으로 자리잡았다. 동아시아 고문헌의 현대적 비평본(critical edition)을 생산하는 학자는 '백문'이라고도 불리는 평문(plain text)에 표점을 부여하여 문장과 문단, 장·절을 나누고 인명·지명·서명 등의 의미 요소들을 식별한다.⁶⁾ 이와 같은 작업을 디지털 세계에서 좀 더 편리하고 깊이있게 수행할 수 있게 해주는 것이 바로 태그(tag)를 활용하는 마크업 언어이다. XML은 원시 자료를 "있는 그대로의 모습"에 가까운 형태로 정보화할 수 있다.⁷⁾ 나아가 텍스트의 구조 및 인명·지명 등의 요소를 식별할 수 있을 뿐 아니라 깊은 해석의 층위까지도 텍스트 이면에 포함시킬 수 있다. 마크업 언어를 통해 작성된 고문헌 디지털 판본은 단순한 문자열 이상의 정보를 담고 있다. 다층성을 지닌 디지털 판본은 텍스트 이면에 통합된 각종 태그들을 통해 텍스트의 복잡한 의미와 구조를 컴퓨터에 전달할 수 있으며, 웹 상의 수많은 다른 자원들과 연결될 수 있는 무한한 가능성을 열어준다.⁸⁾

그러나 마크업 언어는 디지털 판본의 생산을 가능케 해주는 기술일 뿐, 어떻게 이를 활용하여 고문헌의 디지털 판본을 생산할 것인가는 온전히 인문학자들의 손에 맡겨져 있다. 문장을 끊고 문단을 나누고, 저자를 확정하고 다른 텍스트와의 관계를 결정하는 것은 인문학적 행위이며,⁹⁾ 기본적으로 모든 마크업 행위는 해석적이다.¹⁰⁾ 따라서 마크업 언어를

5) 김현, 2018, 12쪽.

6) 김바로, 2016, 205쪽.

7) 김현, 2006, 191쪽.

8) 웹 상의 문서 내에 데이터와 데이터 사이의 의미 정보를 기술하는 메타데이터가 포함되게 함으로써 정보에 대한 자동화된 접근성을 높이려는 이상을 '시맨틱 웹'이라 한다. 현재의 웹에서는 사람만이 산재한 정보의 의미적 관계를 파악할 수 있지만, 시맨틱 웹에서는 컴퓨터가 그 관계를 추적하여 인간의 지식 증대에 도움을 줄 것 기대된다(김현, 2018, 3쪽).

통해 디지털 판본을 생산하는 일은 결국 해당 텍스트의 의미를 이해할 수 있는 인문학자들에 의해 이뤄져야 한다. 마크업 언어를 통한 디지털 판본의 생산은 고전 문헌 텍스트에서 어떻게 의미 요소들을 판별하여 적절히 태깅할 것인가의 문제로 환원된다. 또 작업 과정에서의 협업과, 산출된 결과물들의 호환을 위해 태깅의 규칙을 협의해야할 필요성도 제기된다.

해외의 언어학 및 문헌정보학 분야에서는 디지털 텍스트 인코딩을 위한 표준으로 TEI가 제안되었다. 텍스트의 전산화에 관심을 가진 학자들이 1987년 처음 논의를 시작한 이래 1990년대부터 국제 공동 프로젝트로서 TEI가 수행되었고, 그 결과물이 “Guidelines for Electronic Text Encoding and Interchange” (TEI P3)라는 보고서로 1994년에 발행되었다.¹¹⁾ 현재는 2007년 11월 2일 발표한 “P5 가이드라인”(the P5 Guidelines)이 적용되고 있다.¹²⁾ 그러나 TEI는 기본적으로 서구 언어학과 문헌정보학 분야에서 시작된 것이기에 역사학, 한문학, 불교학 등의 고문헌 데이터에 내재된 개념 용어나 인물·장소 등의 고유정보들을 처리하는데 미흡함이 지적된다.¹³⁾

국내에서는 김현이 1996년 『조선왕조실록』 전산화를 위한 마크업 규칙을 제안하며, 한국학 자료의 디지털 전환을 선도했다.¹⁴⁾ 2006년에는 한국학 고문헌 자료의 XML 문서 편찬을 위한 전반적 가이드라인을 제시했다. 그는 XML 문서의 요소들을 크게 구조 요소(structural element)와 문중 요소(in-text element)로 구분했다. 구조 요소는 특정 상위 요소와 계층적 구조 관계를 맺는 요소, 문중 요소는 문자 데이터 중간에 삽입되어 특정 어휘 또는 어절의 성격을 알려주는 역할을 수행하는 요소로 정의된다. 문중 요소의 적절한 활용은 텍스트의 가독성을 높여줄 뿐 아니라 관련 자료의 상호 참조와 깊이 있는 내용 분석에도 도움이 된다.¹⁵⁾ 김현이 제시한 XML 전자문서화의 가이드라인은 『조선왕조실록』을 비롯한 각종 한국학 문헌의 디지털 판본 제작에 기준점이 되어 왔다. 어떤 요소들을 식별하여 문중 요소로 삼을 것인가는 대상이 되는 자료의 종류와 내용, 활용 목적에 따라 달라질 수 있다.

따라서 본 연구과제에서는 국사편찬위원회의 ‘디지털 조선왕조실록¹⁶⁾’, 타이완

9) 함형석, 「문헌 독해 방법으로서의 텍스트 인코딩」, 『한국불교학』 86, 2018, 48쪽 각주 19.

10) Lou Burnard, “On the Hermeneutic Implications of Text Encoding,” in *New Media and the Humanities: Research and Applications; Proceedings of the First Seminar Computers, Literature and Philology, Edinburgh, 7-9 September 1998*, ed. Domenico Fiormonte and Jonathan Usher (Oxford: Humanities Computing Unit, University of Oxford, 2001), 44.

11) C. M. Sperberg-McQueen and Lou Burnard, “Guidelines for Electronic Text Encoding and Interchange”(https://www.tei-c.org/Vault/GL/P3/index.htm), Oxford, 1994(Revised, 1999).

12) TEI의 가이드라인 제정에 참여한 옥스포드 대학의 루 버나드(Lou Burnard)는 TEI의 특징으로 다음의 세 가지를 들고 있다. 첫째 텍스트가 구현되는 형식보다 의미를 우선시한다(Meaning before format). 둘째, 특정 소프트웨어에 종속되지 않는다(Software independence). 셋째, 텍스트를 직접 다루는 인문사회학자들에 의해서 표준이 유지된다(Community-driven). Lou Burnard, *What Is the Text Encoding Initiative?: How to Add Intelligent Markup to Digital Resources*, Encyclopédie Numérique (Marseille: OpenEdition Press, 2014), 7-11.

13) 김바로, 「디지털 인문학의 관점에서 본 디지털 불교학」, 『불교학보』 86, 동국대학교 불교문화연구원, 2019, 272-273쪽.

14) 김현, 「<조선왕조실록> 전산화(電算化)를 위한 마크업(MarkUp) 규칙」, 『고문연구』 9, 2006.

15) 김현, 2006, 210쪽.

16) 국사편찬위원회, ‘디지털 조선왕조실록’, https://sillok.history.go.kr/, 2024. 1. 13. 접속.

CBETA(Chinese Buddhist Electronic Texts Association)의 ‘전자불전집성(電子佛典集成)’¹⁷⁾, 한국고전번역원의 ‘한국고전종합DB’¹⁸⁾, 한국학중앙연구원의 ‘디지털 장서각’¹⁹⁾ 등 국내외의 고문헌 데이터 설계를 분석하고 각각의 개선 방향을 모색한다.

또 과제에 참여하는 연구원(수강생)들은 선행 고문헌 디지털화의 사례들을 참조하여 각자 자신이 선정한 텍스트를 디지털 판본으로 변환시키는 경험을 갖게 될 것이다. 이를 통해 마크업 언어의 기술적 숙달은 물론 역사학, 한문학, 불교학 텍스트를 의미론적으로 깊이 있게 인코딩하기 위한 고민을 발전시킬 수 있을 것이다. <고문헌자료정보화연구> 강의는 마크업 언어를 기반으로 하는 설계 및 데이터 구축을 통해, 한국학 고문헌 텍스트를 다양한 방식으로 분석·활용하며 보존할 수 있는 고문헌 정보화 모델의 개발을 목표로 하고 있다.

2 선행 연구/수업과의 차별성

2.1. 연구적 측면에서의 차별성

현대 인문학 연구는 정보기술의 도움을 받아 연구자료의 수집부터 연구 결과 공개까지 모든 연구의 전 과정이 컴퓨터에서 이뤄지고 있다. 즉, 인문학 연구자들은 의도적이든 비의도적이든 적절한 디지털 기술을 활용하여 연구의 효율성과 본인의 역량을 높일 수 있는 디지털화 된 연구 환경에 놓이게 되었다. 인문학 연구자에게 요구되는 디지털 기술 관련 지식과 디지털 리터러시 능력의 중요성은 지속적으로 증가하고 있다.

미래의 디지털 기술을 활용한 창의적 인문학 연구활동은 연구 대상 텍스트의 구조와 원문의 내용이 디지털에서도 훼손 없이 유지되는 기본적인 디지털화 뿐만이 아니라, 텍스트에 대한 인문학자의 다양한 판단과 사유를 컴퓨터가 온전히 이해할 수 있는 형태로 설계-구축된 인문 데이터를 필요로 한다.

아날로그 고문헌 텍스트의 인명, 지명, 서명 등의 기본적인 의미 요소부터 텍스트에 대한 다양한 연구자들의 사유까지 모든 관련 정보를 모두 데이터로 만들 수 있다면, 1차적으로는 연구자들의 연구 효율이 상승할 뿐만이 아니라, 2차적으로는 컴퓨터에 의해서 연구자의 의도에 부합하는 데이터 분석을 빠르게 수행할 수 있다. 또한, 인문 데이터의 공유를 통해서 연구 성과를 수정, 보완, 집대성 할 수 있다. 이것이 바로 인문정보학과에서 추구하는 인문 데이터이며, 이는 디지털인문학의 필수불가결한 영역이다.

연구과제의 연구 부문에서 연구원(수강생)들은 개인 또는 2인 1조로 팀을 이루어 국내외의 고문헌 전산화 사례들을 분석하고 각각의 개선 방향을 모색하고, 최종적으로 산출된

17) CBETA 財團法人佛教電子佛典基金會, ‘CBETA 電子佛典集成 2018’, <https://www.cbeta.org/cd/index.php>, 2024. 1. 10. 접속.

18) 한국고전번역원, ‘한국고전종합DB’, <https://db.itkc.or.kr/>, 2024. 1. 10. 접속.

19) 한국학중앙연구원, ‘디지털 장서각’, <https://jsg.aks.ac.kr/>, 2024. 1. 10. 접속.

인문학적 성과를 공동의 성과로 종합함으로써 역사학, 한문학, 불교학 등 한국학 문헌 데이터에 내재된 개념 용어나 인물·장소 등의 고유정보들을 처리할 수 있는 고문헌 정보화 모델 개발의 선행 연구를 수행함으로써 현대 인문학 연구자에게 요구되는 역량을 성취하는 동시에 기존의 인문학과 차별되는 “소통”과 “협업”의 연구 덕목을 실천할 수 있다.

2.2. 교육적 측면에서의 차별성

역사학·한문학·불교학 분야에서 중요 연구자료로 활용되는 고문헌은 공통적으로 인명·지명·개념어 등의 의미요소가 함의되어 있다. 이러한 고문헌 텍스트를 마크업 언어를 이용하여 이뤄지는 텍스트 인코딩 작업은 단순 타이핑이 아닌 해당 문헌의 깊이 있는 이해와 지식을 가지고 진행되어야 할 연구 활동이다. 본 연구과제의 교육을 통해 역사학·한문학·불교학 분야에서 중요 연구자료로 활용되는 고문헌 텍스트를 디지털 판본으로 전환하는데 필요한 적절한 지식과 경험을 학습하여 인문학자가 고문헌 자료의 디지털화 작업에 주체적인 역할을 할 수 있는 능력을 키울 수 있을 것이다.

더불어 교수의 지도하에 일방적으로 전공지식을 전달받았던 기존의 강의식 수업이 아닌 연구원(수강생)들은 그 성과를 한국학 연구자와 대중을 위한 디지털인문학 교육 자료로 제작하는 참여 방식을 통해 전공지식을 능동적으로 배우고 익힐 수 있다. 또한, 연구원(수강생)들은 인문학적 지식을 디지털 언어로 표현하는데 필요한 능력 습득과 협업을 위한 경험을 토대로 본인들의 연구영역에서 창조적인 인문학 연구, 교육 활용을 수행할 것이다.

2.3. 응용적 측면에서의 차별성

한국 고문헌은 여러 인문학 분야의 연구자들에게 중요한 자료인 동시에 인접 학문의 종사자에서부터 작가, 언론인, 대중들에게 다양한 방식으로 활용될 수 있는 원천이다. 이것은 조선시대 500여년 동안 일어난 정치, 경제, 사회, 문화 제분의 사건과 그것에 대한 당대 지식인들의 논의를 상세하게 기록한 조선왕조실록의 디지털화 사례를 통하여 증명되고 있다. 해당 고문헌의 디지털화는 방대한 분량의 고문헌 자료를 읽는 것에만 많은 시간을 할애하였던 조선시대사 연구자나 다른 인접 분야의 연구자에게 다양한 방식으로 해당 문헌에 접근할 수 있는 길을 열었다. 방송작가, 언론인 등 콘텐츠 제작자에게는 풍부한 창작 소재에 직접 접근함으로써 자유롭고 다양하게 응용할 수 있는 기회를 제공하기 때문이다.

본 연구과제에서는 보다 더 다양한 계층의 한국학 고문헌 텍스트 활용 유도를 목적으로 다양한 방식으로 분석·활용하며 보존할 수 있는 고문헌 정보화 모델 개발을 수행하고자 한다. 인문학적 성과물을 위키독스(WikiDocs)와 유튜브(Youtube)를 통해 불특정 다수에게 공개함으로써 대중에게는 한국학 고문헌 자체에 대한 관심과 가치를 전달하고 다른 인문학자에게는 디지털 환경에서의 고문헌 텍스트의 활용 방법론을 전달함으로써 인문학

적 지식과 정보가 정보를 소비하는 계층의 요구에 맞게 활용될 수 있다.

3 연구 내용 및 방법

3.1. 연구-교육 연계과제 결과물

본 연구과제에서는 2024년 1학기 인문정보학과에 개설되는 <고문헌자료정보화연구> 강의와 연계하여, 인문지식 데이터베이스 설계·구축의 대상이 되는 고문헌의 설계 사례분석에 관한 연구문서를 편찬하고, 연구문서를 기반으로 리뷰 콘텐츠를 제작하여 이를 디지털 환경에서 공유한다. 대표적인 선행 설계사례 리뷰 문서는 위키독스(WikiDocs) 플랫폼을 통해서 교재로 편찬하며, 선행 설계사례 리뷰 동영상은 15분 내외의 영상으로 이루어진 6개 강의로 구성하여 유튜브(Youtube)를 통해 공개한다.

표1) 연구교육 연계과제 결과물 리스트

분류	연구-교육 연계과제 결과물	교수	연구원(수강생)
교재	XML 설계 입문 - XML 개념 및 국제 표준 탐색 - XML 설계 방법론	책임	
	전자 텍스트 기술언어 TEI - TEI 개념 및 국제 가이드라인 탐색 - 활용 사례와 연구 동향	강의 및 실습 지도	연구원 1, 2
	【국사편찬위원회 고문헌 설계】 - 디지털 조선왕조실록 소개 및 설계 검토 - 활용 사례와 연구 동향		연구원 3, 4
	【해외 고문헌 설계】 - CBETA 소개 및 설계 검토 - 활용 사례와 연구 동향		연구원 5, 1
	【한국고전번역원 고문헌 설계】 - 한국고전종합 DB 소개 및 설계 검토 - 활용 사례와 연구 동향		연구원 2, 3
	【한국학중앙연구원 고문헌 설계】 - 디지털 장서각 소개 및 설계 검토 - 활용 사례와 연구 동향		연구원 4, 5
동영상	XML 설계 입문	책임	
	전자 텍스트 기술언어 TEI	강의	연구원 1, 2
	【국사편찬위원회 고문헌 설계: 디지털 조선왕조실록】	및 실습 지도	연구원 3, 4
	【해외 고문헌 설계: CBETA(전자불전)】		연구원 5, 1
	【한국고전번역원 고문헌 설계: 한국고전종합 DB】		연구원 2, 3
	【한국학중앙연구원 고문헌 설계: 디지털 장서각】		연구원 4, 5

3.2. 2024년도 1학기 인문정보학 <고문헌자료 정보화연구>강의

- <고문헌자료 정보화연구> 강의는 교수의 강의와 수강생의 실습으로 이루어진다.
- <고문헌자료 정보화연구> 강의는 각 수강생의 연구 주제에 맞는 고문헌 자료를 정하고 그 자료를 데이터로 변환하는 데 필요한 기술을 중심으로 교육한다.
- <고문헌자료 정보화연구> 강의는 실무 프로젝트 수행의 관점에서 사업의 기획과 예산 운영부터 실무 자료의 데이터 설계·가공·변환 및 교육 콘텐츠의 편찬까지 일련의 과정을 수강생들이 모두 경험할 수 있게 하여 단순 지식 습득에 머물지 않고 실무 능력까지 배양할 수 있도록 한다.

표2) <고문헌자료 정보화연구> 강의계획서

교과목명	고문헌자료 정보화연구
교과목표	고문헌 자료의 여러 가지 형식과 내용을 파악하고 각 자료의 내용과 구조를 컴퓨터가 인식할 수 있는 언어로 기술하는 방법을 학습한다. 각 데이터가 공공데이터로 통용될 수 있도록 표준 기술인 TEI를 준수하여 XML데이터 변환을 실습하고, 디지털 자료의 이해와 디지털 리터러시 능력을 증진시킨다. 수강생들은 이 강좌를 통해 여러 종류의 고문헌 자료를 디지털 형식의 텍스트로 변환하고 이를 활용하는 지식을 습득할 수 있다.
강의일자	강의주제
03 월 05 일	고문헌 자료와 정보화
03 월 12 일	XML(eXtensible Markup Language) 개념 및 역사
03 월 19 일	XML(eXtensible Markup Language) 설계
03 월 26 일	TEI(Text Encoding Initiative) 설계 리뷰 1
04 월 02 일	TEI(Text Encoding Initiative) 설계 리뷰 2
04 월 09 일	전문가 초청 특강 1: 함형석(전남대)
04 월 16 일	선행 사례 리뷰: 국사편찬위원회 디지털조선왕조실록
04 월 23 일	선행 사례 리뷰: 법고문리학원 CBETA(전자불전)
04 월 30 일	선행 사례 리뷰: 한국고전번역원 한국고전종합 DB
05 월 07 일	선행 사례 리뷰 : 한국학중앙연구원 디지털 장서각
05 월 14 일	전문가 초청 특강 2: 윤종웅(윤즈정보개발)
05 월 21 일	현행 고문헌 정보화의 문제점 및 보완 방안 논의
05 월 28 일	고문헌 자료의 XML 설계 및 구축 실습 1
06 월 04 일	고문헌 자료의 XML 설계 및 구축 실습 2
06 월 11 일	실습 과제 결과 발표 및 수행평가

- **고문헌 자료 정보화의 이해:** 실물 자료를 데이터로 변환하는 것은 무엇인지 개념, 목표, 방법에 대한 이론적 토대를 마련한다.
- **디지털 형식 기술 언어:** 텍스트 자료를 디지털로 표현하는 마크업 언어기술인 XML의 개념과 이론을 습득한다. XML의 문법을 확인하여 디지털 자료는 어떻게 기술하는지 활용법을 실습해보고, 실물 자료를 가공 처리할 수 있도록 한다.
- **고문헌 자료 정보화 선행 사례 분석:** 웹에서 제공되는 국내외 여러 고문헌 자료 서비스(디지털 조선왕조실록, CBETA, 고전종합DB, 디지털 장서각 등)를 탐색하여 인문지식 텍스트의 전산화 사례를 살펴본다. 각 서비스의 디지털 자료 형식과 기술 방법을 파악하여 웹에서 보여지는 내용과 시맨틱 데이터로 활용될 수 있는 텍스트를 구분할 수 있도록 한다. 각 서비스의 자료의 특징을 파악하고 국제 표준 기술에 따라 데이터를 제공할 수 있도록 연구해본다.
- **고문헌 자료 전자문서 편찬기술 분석:** 각 연구 주제에 따라 분석할 자료를 선정하고, 자료의 내용과 구조를 파악해본 후 어떤 방식으로 XML 편찬했는지 분석한다.

3.3. 강의와 연구 융합 방법

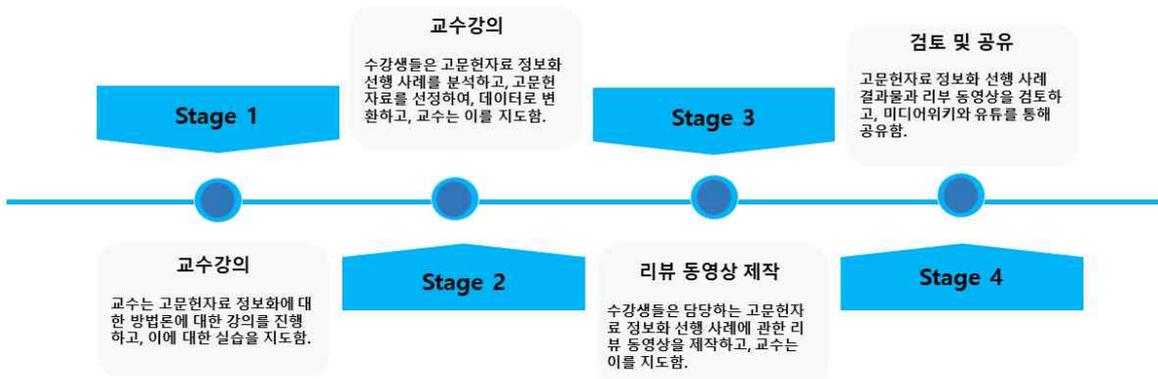


그림2) 연구과제 진행 프로세스

- 교수는 고문헌 자료의 디지털 변환을 위한 기초 이론과 가공 처리에 대한 강의를 수행하고, 연구원(수강생)들의 고문헌 자료 정보화 교육 교재와 동영상 편찬을 지도한다.
- (보조)연구원은 교수의 수강생 지도를 보조하고, 이 경험을 바탕으로 고문헌 자료 정보화 교육 교재와 동영상을 편찬하며, 일정 이상의 기여시 공동저자가 된다.
- 수강생은 교수와 (보조)연구원의 지도를 받아서, 고문헌 자료 정보화 방법론을 숙지하고, 자신의 관심 영역에 대한 고문헌 자료 정보화를 수행한다. 그 중 희망하는 수강생은 고문헌 자료 정보화 교재와 동영상 편찬에 참가할 수 있으며, 일정 이상의 기여시 공동저자가 된다.
- 고문헌 자료 정보화 교육 교재와 동영상 편찬에서 각 항목마다 메인 연구원과 서브연구원을 임명하여, 상호 토의와 검토를 통한 협업 능력을 제고시키고, 결과물의 품질

을 향상하도록 한다.

- 모든 고문헌 자료 정보화 교육 교재와 동영상은 전체 검토 회의를 통해서 상시 점검하고, 최종적으로는 위키독스와 유튜브를 통해서 공유한다.

4 기대효과 및 활용 방안

4.1. 교육적·학문적 의의

√ 디지털 시대 고문헌자료 정보화 방법론 제시

한문 기록으로 남은 과거의 데이터 속에서 우리 문화의 과거와 현재를 잇는 유의미한 사실 관계를 발견하고, 이를 통해 우리의 문화 세계를 통시적으로 볼 수 있는 해안을 얻을 수 있다. 따라서 고문헌자료는 한문학, 역사학, 경학, 고전문학, 동아시아학 등 다양한 분야의 연구대상이 될 수 있으며, 고문헌 속 지식을 발견하고 해석하며 정정하고 재확인하는 작업이 필요하다. 방대한 고문헌 자료를 디지털과 결합함으로써 기존의 연구가 수행하지 못한 자료의 가공과 입력을 완수한다면 새로운 연구의 가능성을 발견하게 될 것이며, 고문헌 자료가 담고 있는 전통적인 인문지식을 배우는 새로운 방법을 제시할 수 있을 것이다.

√ 대학원생 교육방법론 교육 모델 제시

본 연구과제는 가르치는 것은 배우의 절반이며(數學半), 배우면서 가르친다(教學相長, Homines dum docent discunt)는 개념을 바탕으로 역사학·한문학·불교학 등 다양한 분야에서 진행된 고문헌 전산화 사례들을 리뷰하고, XML을 기반으로 동작하는 디지털 텍스트 기술 규약인 TEI도 살펴본다. 이를 바탕으로 고문헌 텍스트를 디지털 판본으로 전환하는데 필요한 적절한 지식과 경험을 학습하고, 그 성과를 위키독스(WikiDocs)와 유튜브(Youtube)를 제작하여 한국학 연구자들을 위해 디지털인문학 교육 자료로 제공한다. 이를 통해서, 연구원(수강생)들의 미래 교수자로서의 교육 능력을 함양할 뿐만 아니라, 대학원생을 대상으로 하는 교육방법론 교육 모델을 제시하는데 의의가 있다.

4.2. 기대되는 수강생의 역량 확대 정도

√ 자기주도 학습역량의 향상

본 연구과제를 수행함에 따라, 연구원(수강생)들은 고문헌 자료를 이해하며 고문헌 자료 데이터화에 대한 전과정에 주도적으로 참여하게 된다. 연구원(수강생)들은 고문헌 자료 정보화의 선행 사례를 리뷰하고, 다양한 분야의 고문헌을 직접 데이터 변환하는 과정을 직접 겪어봄으로써 자기주도 학습 역량을 한층 강화할 수 있다. 일련의 과정은 동영상 제작에 직접 참여함으로써 한층 강화될 수 있다.

√ 인문지식과 정보기술 융합능력의 향상

해당 연구과제를 수행하면 연구원들은 특정 고문헌을 이해하고 선행 연구를 분석하는 능력뿐만 아니라, 디지털 환경에서 다양한 방법을 활용하여 고문헌을 xml 데이터로 변환하고 여러 분야의 문헌을 디지털로 전환하는 경험을 통해 새로운 역량과 성취감을 얻을 것이다. 이는 고문헌 연구를 넘어 디지털 변환과 응용에 대한 탐구로 이어질 것이다. 이 연구과제의 완수를 통해 궁극적으로는 인문지식을 깊이 있게 이해하고, 지식 수용자의 역할을 넘어 창조적인 지식 생산자로서의 역할과 지위를 확립할 수 있을 것으로 기대된다. 이를 통해 새로운 지적인 영역에서의 창의적 기여가 가능해질 것이다.

4.3. 연구결과의 활용 방안

√ 디지털인문학 방법론에 대한 입문서 집필

본 연구과제를 통하여 고문헌 자료의 형식과 내용을 파악하고 고문헌 자료 정보화 선행 사례를 분석할뿐만 아니라 이를 직접 XML 데이터로 변환하는 실습과정에 대한 내용을 담은 디지털 인문학 방법론 입문서를 집필할 것이다. 또한 본 교육 콘텐츠는 디지털 미디어로 제작할 예정이기에, 시간과 공간의 제약을 넘어 학생들이 원하는 시간에 자유롭게 수강 가능하며 향후 인문정보학 전공 학생과 디지털인문학에 관심 있는 대중들에게 입문서로 활용될 것으로 예상된다. 이 자료는 다양한 플랫폼에서도 손쉽게 접근 가능하며, 디지털 분석 방법을 통해 인문 데이터에 접근하는 데 유용한 교양서가 될 것이다.

√ 디지털인문학 유튜브 채널 운영

본 연구과제 수행을 통해 창출된 결과물을 현재 운영중인 디지털인문학 전문 유튜브 채널에 업로드 할 예정이다. 또한, 유튜브를 통해 디지털인문학 교육에 도움이 되는 유용한 교육 콘텐츠를 지속적으로 생산함으로써 디지털인문학도를 양성할 계획이다. 교육 콘텐츠에는 고문헌 자료에 대한 전반적인 이해를 바탕으로 전자 텍스트 기술언어인 TEI에 대해 공부하고, 다양한 고문헌 자료 정보화 선행 사례를 분석하며 앞선 내용을 바탕으로 고문헌 자료를 직접 데이터화하는 과정을 담길 것이다. 이러한 이러한 콘텐츠를 디지털인문학 전문 채널에 제공함으로써 전통적인 방법론 이외의 새로운 접근법을 골몰하던 여러 인문학자들은 보다 용이하게 관련 지식을 습득할 수 있고, 일반 대중 역시 본인의 필요에 맞게 고문헌 자료를 데이터화 하여 디지털 인문학의 단순 소비자가 아닌 생산자로서 기여한다면, 디지털인문학의 확산 및 적용을 위한 새로운 창구로 자리매김할 수 있을 것으로 전망된다.

5 연구자 및 초청전문가의 연구경력

연구책임자:

- 박사학위논문: 「제도와 인사의 관계성 데이터 아카이브 구축과 활용 - 근대 학교 자료(1895~1910)를 중심으로 -」 2017, 한국학중앙연구원 한국학대학원.
- 저서1: 『AI가 내려온다 - 인공 지능 시대의 고전 문학 연구』, 사이언스북스, 2022.
- 저서2: 『디지털 인문학 입문』, 한국외국어대학교지식출판원, 2016.
- 논문1: 「한국역사인물데이터베이스 설계 시론: 홍문록(弘文錄)을 예시로 하여」, 『한국학』, 2023.
- 논문2: 「조선시대 중국어 역학서 데이터베이스 구축 연구」, 『중국학』, 2022.
- 논문3: 「인공지능 시대의 사전과 데이터 - 한국역사인물사전(데이터)을 위한 탐색 -」, 『동양학』, 2021.
- 논문4: 「인공지능 시대를 위한 인문데이터의 역사와 과제」, 『인공지능인문학연구』3, 2019.

초청전문가A:

- 박사학위논문: *Buddhist Critiques of the Veda and Vedic Sacrifice: A Study of Bhāviveka's Mīmāṃsā Chapter of the Madhyamakahrdayakārikā and Tarkajvālā* (2016, University of Michigan)
- 역서1: *Science and Philosophy in the Indian Buddhist Classics* vol. 3: *Philosophical Schools, Wisdom Publications* (with Donald Lopez; 2022)
- 역서2: *An Anthology of East Asian Commentaries on the Nyāyapraveśa*, Seoul: Jogye Order of Korean Buddhism (2017)
- 논문1: "On Being Orthodox Renouncers: the Yuktidīpikā's Establishment of the Sāṅkhya Mode of Life (Sannyāsa) in the Name of the Veda," *Journal of the Royal Asiatic Society*, vol. 33, issue 3, pp. 627-641. (2023)
- 논문2: 「산스크리트 논서(śāstra)문헌군 인코딩 표준 마련을 위한 기초연구」, 『불교학연구』, 대구: 불교학연구회, 제 73호. (2022)
- 논문3: "Bhāviveka's Inclusivism: Discriminating the Feces, Jewels, and Fake Jewels of the Veda," *Philosophy East and West*, vol. 72, no. 2, pp. 315-334. (2022)

- 논문4: 「악인의 저작에서 암흑의 텍스트로 – 미망사학파의 인식의 본유적 타당성(svataḥpramānya) 이론과 불교 베다 비판의 전환점」, 『불교학연구』, 대구: 불교학연구회, 제 65호, pp. 1-27. (2020)
- 논문5: “Manipulating the Memory of Meat-Eating: Reading the Laṅkāvatāra’s Strategy of Introducing Vegetarianism to Buddhism,” *Journal of Indian Philosophy*, vol. 47, no. 1, pp. 133-153. (2019)

초청전문가B

- 박사학위논문: 『역사 자료 텍스트의 전자적 기술에 의한 지식 관계망 구현 연구 : 『通鑑節要』와 『譯註 通鑑節要』를 중심으로』, 한국학중앙연구원 한국학대학원, 2017.
- 경력1: 2005~2007 한국역대인물종합정보시스템 구축 사업 PM
- 경력2: 2007년 조선왕조실록 대국민 온라인서비스 사업 PM
- 경력3: 한국문집총간 DB 구축, 한국족보자료 시스템, 동양고전 DB 구축 등 기타 한국학 관련 DB 구축 사업을 다수 수행
- 연구보고서1: 『東洋古典 漢韓 말뭉치 구축과 활용』, 전통문화연구회, 2023.11

6 참고문헌

논저

- 김바로, 2016, 「역사기록의 전자문서 편찬방법 탐구: 역사요소를 중심으로」, 『열상고전연구』 51, 203-223.
- 김바로, 2019, 「디지털 인문학의 관점에서 본 디지털 불교학」, 『불교학보』 86, 265-288, <https://doi.org/10.18587/bh.2019.3.86.265>
- 김현, 1996, 「<조선왕조실록> 전산화(電算化)를 위한 마크업(MarkUp) 규칙」, 『고문연구』 9, 244-262.
- 김현, 2006, 「고문헌 자료 XML 전자문서 편찬 기술에 관한 연구」, 『古文書研究』 29, 183-230.
- 김현, 2018, 「데이터 기반 漢文古典 研究를 위한 試論」, 전통문화연구회 창립 30주년 기념 학술회의, 전통문화연구회.

- 김현·임영상·김바로, 2016, 『디지털 인문학 입문』, Huebooks.
- 안병학·정우봉·정출현, 2001, 「한국 고전문헌 데이터베이스의 설계·구축 및 응용 방안 연구」, 『민족문화연구』 34, 179-389.
- 엄혜련·김성혁, 1996, 「TEI Independent Header와 MARC의 비교연구」, 『한국정보관리학회 학술대회 논문집』, 한국정보관리학회, 75-78.
- 이남희, 2003, 「디지털 시대의 고문서정리 표준화」, 『고문서연구』 22, 2003, 25-50.
- 이남희, 2017, 「한국사 문헌자료 디지털화의 현황과 과제」, 『열린정신 인문학연구』 18(2), 173-198.
- 이상용, 2003, 「XML을 활용한 고문헌의 원문디지털화 방안에 대한 연구-고문헌을 위한 DTD개발을 중심으로-」, 『한국문헌정보학회지』 37(1), 171-201.
- 이상용, 2003, 「한국 문집을 위한 XML DTD 개발에 대한 연구」, 『서지학연구』 25, 169-208.
- 함형석, 2018, 「문헌 독해 방법으로서의 텍스트 인코딩」, 『한국불교학』, 86, 2018, 35-61.
- Berti, Monica, ed. *Digital Classical Philology: Ancient Greek and Latin in the Digital Revolution. Age of Access?*, volume 10. Berlin, Germany; Boston, Massachusetts: De Gruyter Saur, 2019.
- Burnard, Lou. "On the Hermeneutic Implications of Text Encoding." In *New Media and the Humanities: Research and Applications; Proceedings of the First Seminar Computers, Literature and Philology, Edinburgh, 7-9 September 1998*, edited by Domenico Fiormonte and Jonathan Usher, 39-45. Oxford: Humanities Computing Unit, University of Oxford, 2001.
- . *What Is the Text Encoding Initiative?: How to Add Intelligent Markup to Digital Resources*. Encyclopédie Numérique. Marseille: OpenEdition Press, 2014. <http://doi.org/10.4000/books.oep.426>
- Ciotti, Fabio. "Text Encoding as a Theoretical Language for Text Analysis." In *New Media and the Humanities: Research and Applications; Proceedings of the First Seminar Computers, Literature and Philology, Edinburgh, 7-9 September 1998*, edited by Domenico Fiormonte and Jonathan Usher, 47-54. Oxford: Humanities Computing Unit, University of Oxford, 2001.
- Cummings, James. "The Text Encoding Initiative and the Study of Literature." In *A Companion to Digital Literary Studies*, 451-476. John Wiley & Sons, Ltd, 2013. <https://doi.org/10.1002/9781405177504.ch25>.
- Fiormonte, Domenico, and Jonathan Usher, eds. *New Media and the Humanities: Research and Applications; Proceedings of the First Seminar Computers, Literature and Philology, Edinburgh, 7-9 September 1998*. Oxford: Humanities Computing

Unit, University of Oxford, 2001.

Hockey, Susan. *Electronic Texts in the Humanities: Principles and Practice*. Oxford University Press, 2000. <https://doi.org/10.1093/acprof:oso/9780198711940.001.0001>

Ide, Nancy M., and Jean Véronis, eds. *Text Encoding Initiative: Background and Contexts*. Dordrecht; Boston: Kluwer Academic Publishers, 1995.

Pfannenschmidt, Sarah L., and Tanya E. Clement. "Evaluating Digital Scholarship: Suggestions and Strategies for the Text Encoding Initiative." *Journal of the Text Encoding Initiative*, Issue 7 (November 12, 2014). <https://doi.org/10.4000/jtei.949>.

Siemens, Raymond George, and Susan Schreibman, eds. *A Companion to Digital Literary Studies*. Blackwell Companions to Literature and Culture 50. Malden, MA: Blackwell Pub, 2007.

Sutherland, Kathryn, ed. *Electronic Text: Investigations in Method and Theory*. Oxford: Clarendon Press, 1997.

Tauber, James K. "Character Encoding of Classical Languages." In *Digital Classical Philology: Ancient Greek and Latin in the Digital Revolution*, edited by Monica Bertini, 137–58. De Gruyter Saur, 2019. <https://doi.org/10.1515/9783110599572-009>

Vogeler, Georg. "What Is Text, Really? TEI and Beyond – An Introduction." *Journal of the Text Encoding Initiative*, Issue 14 (March 17, 2021). <https://doi.org/10.4000/jtei.4460>

웹사이트

CBETA 財團法人佛教電子佛典基金會, 'CBETA 電子佛典集成 2018', <https://www.cbeta.org/>

국사편찬위원회, '디지털조선왕조실록', <https://sillok.history.go.kr/>

한국고전번역원, '한국고전종합DB', <https://db.itkc.or.kr/>

한국학중앙연구원, '디지털 장서각', <https://jsg.aks.ac.kr/>

Ⅱ. 연구 추진 계획

연 월 일	연구수행 내용 및 연구 역할	비 고
2024. 3.	<p>연구수행 내용: 고문헌 자료 정보화의 학습 및 교육 콘텐츠 제작을 위한 주제와 자료 조사.</p> <p>연구 역할: 연구원(수강생)들은 고문헌 자료 정보화에 대한 이해를 바탕으로 인문학 텍스트 전산화의 역사와 한국학 고문헌 자료의 디지털화에 유용한 선행 사례를 탐색하고 관련 자료를 광범위하게 조사하여 논의한다.</p>	
4.	<p>연구수행 내용: 국사편찬위원회 고문헌 데이터 설계에 관한 리뷰 문서 작성 및 교육 콘텐츠 편찬.</p> <p>연구 역할: 연구원(수강생)들은 각자 맡은 연구수행을 위해 개인 또는 팀 단위로 나뉘어 디지털 조선왕조실록 소개 및 XML 문서 검토, 역사 사료의 구조적 특성, 활용 및 연구 동향을 분석 및 정리한다. 이를 바탕으로 역사 사료 관련 XML에 관한 상세한 정보를 제공하는 리뷰 문서와 쉬운 이해를 돕는 교육 동영상을 제작한다.</p>	
5.	<p>연구수행 내용: 불경 데이터 설계에 관한 리뷰 문서 작성 및 교육 콘텐츠 편찬.</p> <p>연구 역할: 연구원(수강생)들은 각자 맡은 연구수행을 위해 개인 또는 팀 단위로 나뉘어 CBETA 소개 및 TEI 문서 검토, 불교문헌의 구조적 특성, 활용사례 및 연구 동향을 분석, 번역, 정리한다. 이를 바탕으로 불교문헌에 관한 상세한 정보를 제공하는 리뷰 문서와 쉬운 이해를 돕는 교육 동영상을 제작한다.</p>	

6.	<p>연구수행 내용: 한국고전번역원과 한국학중앙연구원의 고문헌 데이터 설계에 관한 리뷰 문서 작성 및 교육 콘텐츠 편찬.</p> <p>연구 역할: 연구원(수강생)들은 각자 맡은 연구수행을 위해 개인 또는 팀 단위로 나뉘어 한국고전DB와 디지털 장서각 소개 및 XML 문서 검토, 고문헌 자료의 종류 및 구조적 특성 활용 및 연구 동향을 분석 및 정리한다. 이를 바탕으로 고문헌 관련 XML에 관한 상세한 정보를 제공하는 리뷰 문서와 쉬운 이해를 돕는 교육 동영상 제작한다.</p>	
7.	<p>연구수행 내용: 기존 XML 문서의 문제점 및 보완 방안 작성 및 교육 콘텐츠 편찬.</p> <p>연구 역할: 연구원(수강생)들은 각자 맡은 연구수행을 위해 개인 또는 팀 단위로 나뉘어 기존 XML 문서의 문제점 및 보완 방안을 탐색하고, 해당 내용을 바탕으로 고문헌 텍스트를 선정하여 XML 편집도구를 활용한 실습을 진행한다. 이를 바탕으로 한국학 고문헌 XML 설계 실습에 관한 상세한 정보를 제공하는 리뷰 문서와 쉬운 이해를 돕는 교육 동영상을 제작한다.</p>	
8.	<p>연구수행 내용: 연구 과제 결과물 최종 검토 및 제출.</p> <p>연구 역할: 연구원(수강생)들은 자신들이 제작한 연구 과제 결과물의 최종본을 제출하게 되며, 이 때 연구책임자는 최종결과물의 검토 및 감수를 진행하여 제출한다.</p>	

Ⅲ. 연구비

(단위 : 원)

항목		예산액	산출근거	비고
비목	세목			
인건비	학생 인건비	15,000,000(원)	○ 석·박사과정: 50만원 X 5명 X 6개월 = 1천5백만원	
직접비	장비 및 재료비			
	학 술 연구비	1,000,000(원)	○ 전문가 초빙: 50만원 X 2명 = 100만원	
	학술활동수당	<비해당>		
간접비		<비해당>		
부가가치세		<비해당>		
합계		16,000,000(원)		

연구결과물 처리방식 (○표)	단행본출판	○	출판예정일	2026년 08월
	기타 결과물	온라인교재(0), 동영상(0)		
연구결과물의 양		선행 마크업 설계 리뷰 문서 6건 교육 동영상(15분 분량) 6건		
결과보고일		2024년 8월 30일		
<p>위와 같이 2024년도 연구계획서를 제출하며, 연구협약서 내용 및 관계 규정, 연구윤리규정을 성실히 이행할 것을 약속합니다.</p> <p style="text-align: center;">2024년 01월 31일</p> <p style="text-align: right;">연구책임자 김 바로 (인) </p> <p>한국학중앙연구원장 귀하</p>				