

A Study on the Data Design for the Construction of Database the Korean Association for Buddhist Studies based on Dongguk University's Combined Buddhist Canon

동국대학교 통합대장경 기반 한국 불교학 통합 DB 구축을 위한 데이터 설계 연구

정송이 (Song-yi Jung), 한국학중앙연구원

김바로 (Ba-ro Kim), 한국학중앙연구원

본 연구는 시맨틱 웹(Semantic Web) 개념을¹ 활용하여, 동국대학교 불교기록문화유산 아카이브(ABC 아카이브, Archives of Buddhist Culture)의 통합대장경 데이터를 대상으로, 다층적이고 풍부한 인문학적 지식정보를 제공할 수 있도록 하는 불교학 데이터모델 설계를 제안하고자 한다.

이를 위해서, 동국대학교 ABC 아카이브 통합대장경, 대만 중화전자불전협회(일명 CBETA, 財團法人佛教電子佛典基金會, Chinese Buddhist Electronic Text Association), 대만 불학규범자료고(佛學規範資料庫, Buddhist Studies Authority Database Project) 등의 유관 불교학 데이터 설계를 살펴보고 전세계 불교학 데이터와 호환성을 가지면서 한국 불교학만의 고유성을 반영할 수 있는 데이터 온톨로지(Ontology) 설계를 수행한다.

동국대학교 통합대장경은 현재 총 1514건의 자료가 구축되어 있으며, 한글대장경과 고려대장경 원문 이미지와 원문 텍스트와 한국어 텍스트를 열람하고 검색할 수 있다.² 대장경의 내용의 맥락과 구조를 그대로 유지하고 한국어 번역문과 함께 제공하고 있지만³, 인명·지명·시간 요소에 함의된 다층적인 인문학적 지식정보를 연계할 수 있는 설계는 보이지 않는다.

¹ 시맨틱 웹의 개념: https://ko.wikipedia.org/wiki/시맨틱_웹 (검색일자: 2024.03.20.)

² 동국대학교 불교기록문화유산아카이브통합대장경소개: <https://kabc.dongguk.edu/etc/intro?type=02> (검색일자: 2024.03.20.)

³ 이재수, 「한글대장경 전산화의 미래 과제와 활용 방안」, 『전자불전』 제14집(동국대학교 전자불전문화콘텐츠연구소, 2012), 112쪽.

대만 CBETA는 불교경전의 새로운 학술 디지털판 제작을 추진하고자, 1998년을 시작으로 『大藏經』, 『大正新脩大藏經』, 『歷代藏經補輯』부터 근래에는 『太虛大師全書』 등 불교 고문헌자료가 구축되고 있다.⁴ 더불어 메타데이터를 통하여 서지정보를 제공하고 XML·TEI(Text Encoding Initiative)를 활용하여⁵ 원문 텍스트를 문장 단위로 전달하는 동시에 최대한 원형을 보존하며, 원문 텍스트에 출현하는 인명·지명·시간 용어의 핵심정보를 마크업(markup)하고 있다.

불학규범자료고는 CBETA 원문 데이터와 대만 법고불교대학(法鼓佛教大學, Dharma Drum Buddhist College)의 다양한 프로젝트에서 얻은 시간, 인명, 지명, 불경목록 데이터를 RDB(Relational Database)를 기반으로 통합·구축된 기초데이터베이스이다.⁶

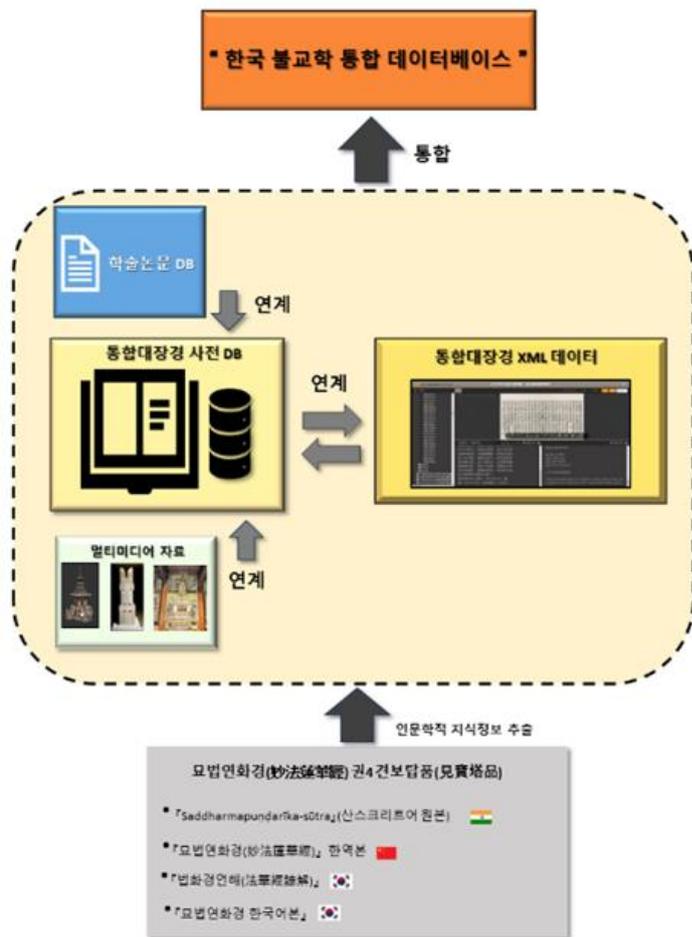
CBETA는 불학규범자료고 데이터와의 연계를 통하여 불경에 대한 추가적인 정보를 제공하고 있지만, 불경 원문에 대한 이해도를 높일 수 있는 원문에 나오는 특정 키워드와 관계된 멀티미디어 데이터 미제공은 아쉬운 부분이다.⁷

⁴ CBETA는 불교경전 전산화에 종사하는 개인과 단체가 모여서 시작되었다. 1997년 10월 샤오첸궈(蕭鎮國)가 CCCII(Chinese Character Code for Information Interchange, 정보교환을 위한 한자코드)로 인코딩된 25권 분량의 <대정장> 전자원고 제공을 시작으로 국립대만대학 불교연구센터에서 대규모 장경 전산화 작업을 위한 '25T'를 결성하였고, 1998년 형칭(恆淸)스님의 노력으로 북미 인순재단과 대만 법고산 중화불학연구소의 대대적인 지원과 후원을 받으면서 1998년 2월 15일에 정식으로 설립되었다. 採訪整理, 「締造佛教電子出版的新紀元 - 中華電子佛典協會-」, 『佛教圖書館館訊』 제24기(香光尼眾佛學院圖書館, 1989): <http://www.gaya.org.tw/journal/m24/24-publish.htm> (검색일자:2024.03.20); CBETA 緣起: https://cbetaonline.dila.edu.tw/doc/en/05-03_origin.php (검색일자:2024.03.20);CBETA 基金會成立簡介:<https://www.cbeta.org/node/6384>(검색일자:2024.03.20);CBETA 目標: <https://www.cbeta.org/intro/mission.php> (검색일자: 2024.03.20)

⁵ XML버전 형식 설명(XML版格式說明): <https://www.cbeta.org/format/xml.php> (검색일자: 2024.03.20);파일코딩교환형식소개(文件編碼交換格式簡介): https://metadata.teldap.tw/standard/rarebook/TEI921224/index_c.htm (검색일자: 2024.03.20)

⁶ 불학규범자료고: <https://authority.dila.edu.tw/> (검색일자: 2024.03.20.)

⁷ 김바로, 「디지털 인문학의 관점에서 본 디지털 불교학」, 『불교학보』 제86집(동국대학교 불교문화연구원, 2019), 281쪽.



현 시점에서 가장 방대한 불교학 데이터를 구축한 CBETA와 불학규범자료고의 데이터 구조를 참조하여 한국 전자 불전에 최적화될 수 있는 데이터 설계를 제안하고자 한다. 통합대장경 원문 텍스트에서 인명·지명·시간 정보를 추출하여 해당 데이터를 통합대장경 사전 데이터베이스로 통합·구축하여 통합대장경 원문 및 한국어 텍스트와 연계한다.

또한 원전을 저본(底本)으로 조선시대 언해본, 근현대시대 한국어 번역본, 현대 한국어 번역본을 연계하여 한국 불교학의 특징을 반영하고자 한다. 더 나아가 멀티미디어 시대에 부합하는 불교학 통합 데이터베이스를 목표로 이미지, 3D, 동영상과 같은 다양한 형태의 시각자료를 상호 연계하고자 한다.

부처의 말씀을 담은 불교경전은 불교고대부터 팔리어 원전을 저본으로 산스크리트어본, 한역본 등 다양한 언어로 번역되어 세계 각국으로 전파되었다. 이것은 불교사상을 넓게 전파하기 위한 것으로 그 과정에서 근본 경전을 바탕으로 학생들의 학문적 저술활동과 불상이나 불화를 조성하는 행위가 오랜 시간 동안 이루어졌다. 이러한 맥락에서 대장경 텍스트에 내포된 여러 불교학 정보를 추출하여

통합 DB로 구축하는 동시에 멀티미디어 자료들과 연계하는 것은 디지털 시대의 불교학 연구라고 할 수 있다.

이러한 동국대학교 통합대장경 기반 불교학 통합 데이터베이스 구축 연구는 한국 불교학 연구자들의 연구에 다양하게 활용할 수 있는 원천자료를 제공함과 동시에 불경을 바탕으로 불교사상을 학습하고 싶은 대중에게 원문에 대한 접근성을 용이하게 함으로써 불교 대중화에 기여할 것으로 기대된다.