

온톨로지 기반 한국어 학습자료

안내 사이트 개발*



한교도서관팀

이세희(팀장), 1911098, tpgml0502@naver.com

박서정, 1911088, 0305sjpark@naver.com

이수민, 1911068, sudl000925@naver.com

이예지, 1911008, mary.lee07164@gmail.com

목차

1. 프로젝트 개요
 - 1.1 프로젝트 정의
 - 1.2 프로젝트 배경 및 목표
 - 1.3 결과물의 활용가치
 - 1.4 제한점
 - 1.5 프로젝트 수행방법

2. 프로젝트 결과물
 - 2.1 프로젝트 결과물 설명
 - 2.2 구조 및 기능
 - 2.2.1 표준체계 : 온톨로지 구축
 - 2.2.2 적용예시 : 홈페이지 구축
 - 2.3 프로젝트 결과물
 - 2.4 장단점
 - 2.5 기대효과
 - 2.6 관련 기술 및 개발방법
 - 2.6.1 관련 기술
 - 2.6.2 개발 도구

3. 프로젝트 추진 체계 및 일정

4. 참고문헌 및 참고한 웹 사이트

표 목차

- 표 1. 온톨로지 표
- 표 2. 클래스 조직 표
- 표 3. 조원 조직도
- 표 4. 프로젝트 진행 일정표

그림 목차

- 그림 1. 실제 온톨로지 구축 예시 1
- 그림 2. 실제 온톨로지 구축 예시 2
- 그림 3. 실제 온톨로지 구축 예시 3
- 그림 4. 실제 온톨로지 구축 예시 4
- 그림 5. 한교 도서관 로고
- 그림 6. 한교 도서관 사이트 첫 시작 페이지
- 그림 7. 한교 도서관 사이트 <Book> 페이지
- 그림 8. 한교 도서관 사이트 <TextBook> 페이지
- 그림 9. 도서의 하위 페이지 1
- 그림 10. 도서의 하위 페이지 2
- 그림 11. 도서의 하위 페이지 3
- 그림 12. 도서의 하위 페이지 4
- 그림 13. 도서의 상세 페이지 1
- 그림 14. 도서의 상세 페이지 2
- 그림 15. 도서의 상세 페이지 3
- 그림 16. 조원 역할분담

1. 프로젝트 개요

1.1 프로젝트 정의

'한교도서관'은 '한국어 교육 도서관'이라는 의미로, 도서관에서의 한국어 교육을 지원하고자 붙인 팀명이다. 여기서 '도서관'의 의미는 실제 책을 열람할 수 있는 도서관에 국한하지 아니하고 특정 자료들을 다양한 기준에 따라 소개하는 곳을 말한다. 일반적인 도서관과의 가장 큰 차이로는 원하는 방식으로 자료들을 분류해 브라우징할 수 있다는 것, 관련된 자료들 간 연결해주어 이용자의 선택 환경을 확장시켜준다는 것이다. 이렇듯 자료 소개의 표준을 제공해 기관별 유형에 구애받지 않고 유동적으로 사용할 수 있는 도서관이라는 폭넓은 개념으로 쓰인다.

'온톨로지 기반 한국어 학습자료 안내 사이트 개발' 프로젝트는 그러한 도서관에서 여러 기관의 다양한 한국어 학습자료에 대한 정보를 하나로 모아 제공해주기 위해 한국어 자료 분류의 통합된 체계를 마련하고자 추진한 활동이다. 선정한 몇 가지 분류 기준에 따라 온톨로지를 이용해 표준 체계를 만들고, 이를 기반으로 간단한 홈페이지 예시 화면도 제작해 보여줌으로써 프로젝트를 구체화 하였다.

1.2 프로젝트 배경 및 목표

현재 한국어 학습자료를 출판하는 공식 기관 중 하나인 세종학당은 물론, 다양한 대학과 교육청 사이트에서는 각각 다른 체계와 방식을 사용하여 한국어 학습자료의 정보를 전달하고 있다. '국립국어원'의 경우, 제공 대상을 국내 유아·학령기·이주민·성인, 국외 유아·성인으로 구분하고, 도서명·학습자단계·제공처·등록일·기능·영역·유형·내용을 도서 정보로 제공하고 있다. 또한, 기능·영역·유형·학습자단계 네 가지를 상세검색의 기준으로 한다. '세종학당'의 경우, 크게 교재·맞보기

강의·학습자료·모바일앱으로구분해도서명·난이도·저작권·등록일·교육과정·제공자료·내용소개·목차·구
매안내를 도서정보로 보여주며, 상세검색 기능은 없다. '서울대학교 한국어교육센터'에서는
시리즈로만 구분해 간단한 도서명과 급수 내용만 소개 하고 상세검색은 제공하지 않는다.

'연세대학교 한국어교사연수소'도 시리즈로 구분하고 있으며 도서명·급수·기능·예시 스크립트를
도서정보로 소개하며 마찬가지로 상세검색은 지원하지 않는다. '서강대학교 한국어교육원'에서도
시리즈로 구분, 도서명·교재구성표·답안지·교재 미리보기·교재구입 안내·cd/mp3 파일·mp3 QR 코드의
도서정보를 담고 있으며, 상세검색을 지원하지 않는다. 다문화센터 홈페이지에서도 한국어
학습자료에 대한 소개나 안내가 누락되어 있는 곳이 대부분이다.

이처럼 같은 한국어 학습자료임에도 소개하는 기관에 따라 제공되는 자료 구분, 정보 및 제공
형식은 물론, 정보 유형의 명칭 또한 다르다는 것을 확인할 수 있다. 한편, 대부분의 다문화
센터에서는 프로그램에 대한 소개만 있을 뿐, 참고할 수 있는 한국어 학습자료에 대한 정보가 없기
때문에 일정 및 사정상 한국어 교실이나 학원에 다닐 수 없는 학습자의 경우에는 직접 한국어 교재
및 학습자료를 찾아보고 공부해야하는 불편한 상황으로 이어진다.

'한교도서관'에서는 이러한 두 가지 문제 상황을 바탕으로 정형화된 한국어 학습자료 소개 기준의
필요성을 인식하였다. 따라서 이 프로젝트에서는 기관 및 사이트마다 현재 통일되어 있지 않은
한국어 학습자료의 제공 방식 표준화 및 배포를 목적으로 하며, 표준을 토대로 홈페이지를 직접
구축하여 사용자에게 공통된 지표로써 보여줄 수 있도록 한다. 최종적으로 프로젝트의 기준이 될
온톨로지를 구축한 후 한국어 학습자료를 소개하는 내용이 담긴 사이트인 국립국어원, 세종학당, 각
대학교의 한국어 교육센터, 다문화센터 사이트에서 자유롭게 사용할 수 있도록 함을 목표로 한다.

1.3 결과물의 활용가치

프로젝트 결과물인 온톨로지는 프로젝트에서 한국어 학습자료의 분류 기준으로 삼은 요소들을
보여주고, 이해를 돕고자 하는 용도로 사용된다.

이해를 돕고자 Wix 프로그램을 이용하여 구축한 온톨로지 기반의 실제 홈페이지는 온톨로지가

어떤 식으로 사용될 수 있는지 알 수 있는 예시로 활용되며, 더 나아가 홈페이지 자체가 한국어 학습자료 소개 페이지의 초안 역할을 하는 것도 가능하다.

본 프로젝트는 한국어 학습자료를 넘어 영어책, 혹은 각 대학에서 사용하는 교재와 같이 보여주고자 하는 단일 주제의 자료들을 적용시켜 표준화된 자료 제공 기준을 제시할 수 있도록 응용하는 것도 가능할 것으로 예상된다.

1.4 제한점

1) 전문성

웍스에서는 동적페이지를 사용해 무제한으로 사이트 페이지를 생성할 수 있다. 그러나 기술적인 문제로 동적페이지로 만들지 못하고 정적페이지로 만들었는데, 정적페이지는 최대 100 개로 제한되어 조금 더 많은 정보를 담지 못했다. 그렇지만 인스턴스 부분의 양이 적을 뿐 홈페이지의 구조와 형식을 구성하는 다른 페이지는 문제없이 구축하였다.

2) 윤리성

홈페이지를 구축할 때 사용되는 도서 사진이나 영상자료 등과 관련된 저작권 문제가 발생할 수 있다. 이를 해결하기 위해 저작물의 이용허가 범위를 확인하고 그 범주 내에서 이용하거나 저작권자에게 사용허가를 받은 후 이용, 혹은 출처 표기를 하여 이용하고자 한다.

3) 미학

홈페이지 예시안은 홈페이지 제작을 도와주는 사이트인 'Wix'를 통해 구축하였다. 실제 홈페이지를 전문적으로 구축한 것은 아니며, Wix 에서 제공하는 도구만으로 구성해야 했다. 또한, 그마저도

지식의 한계로 사용하지 못한 도구들이 있어서 미적으로 최상의 결과물을 끌어낼 수는 없었다. 하지만, Wix 를 통해 기능 및 디자인 측면에서 최대한 다른 사이트들과 크게 차이가 나지 않는 모양새를 갖추도록 하였고. 다양한 국적의 사람들이 이용 대상이므로, 가능한 직관적인 화면 배치와 이해하기 쉬운 아이콘 등을 배치해 이용자의 편의성에 초점을 맞췄다.

4) 사회에 미치는 영향

한국어 학습자료 분류 프로젝트의 표준이 실제 여러 기관의 사이트에 적용될 경우, 갑작스러운 홈페이지 변화로 기존 이용자가 적응하지 못하거나 이용에 어려움을 겪는 상황이 발생할 수 있다. 이러한 문제점을 가정하여 최대한 간편한 검색과 번역이 동원되어 정확한 번역으로 이어질 수 있도록 직관적인 단어들로 구성한다.

5) 제한된 정보와 광범위한 자료

공식 제공처에서 영역이나 기능 부분의 정보가 표기되어있지 않은 경우가 상당하여 수집하는데 어려움이 있다. 이런 경우를 대비하여 책소개에서 유추해 기입하거나 온톨로지에 “알 수 없음” 인스턴스를 제작하여 이에 포함시켰다. 또한, 분류 대상의 유형이 광범위하여 교재, 익힘책, 동영상파일, 음성파일 등과 같이 가장 대표적인 유형으로만 한정지어 구축하였다.

6) 신뢰성

해당 표준 어휘를 다른 개발자가 사용할 때 본 프로젝트의 의도와 다르게 해석되어 사용될 수 있다. 이를 방지하기 위하여 최대한 메타데이터 표준을 사용하고자 하였고 온톨로지에는 모든 클래스 및 프로퍼티에 `rdfs:label` 과 `rdfs:comment` 를 통해 용어를 더 쉽게 이해할 수 있는 형태 및 용어의 의미, 작성 방식이나 유의사항 등에 대한 정보를 제공하였다.

1.5 프로젝트 수행방법

1) 한국어 교재 소개사이트 및 인스턴스 조사

먼저 한국어 교재를 소개하고 있는 사이트는 어느 곳인지 알아보았으며, 그 결과 대표적으로 국립국어원, 세종학당, 각 대학교의 한국어 교육센터, 그리고 다문화센터 등이 있음을 확인하였다. 이 과정 중 국립국어원 제공의 연구보고서인 '2020 국내외 한국어 교재 사용 현황 조사'에서 한국어 학습자료 교재 종류와 국내외의 한국어 교재 사용 기관에 대한 많은 정보를 얻을 수 있었다. 주로 국내는 KSL(Korean as a Second Language: 제 2 언어로서의 한국어), 사회통합프로그램, 다문화가족지원센터에서 교재를 이용하고 국외는 세종학당, 한글학교, 각 대학교재를 이용한다는 정보를 이용해 수월하게 사이트와 교재를 조사할 수 있었다.

위 사이트를 하나하나 둘러보며 어떠한 정보를 어떤 형식과 정보명으로 제공하고 있는지, 해당 사이트에서 준수하고 있는 자료 제공 방식과 포함 요소들을 일차적으로 조사하였다. 이후 수집한 정보 가운데 온톨로지 표준체계로 사용할 만한 내용을 걸러내어 분산되어 있는 한국어 학습자료가 어떤 식으로 분류되면 좋을지 그 기준에 대해 논의하고 정리하였다.

온톨로지를 구축하기 위해서는 각 계층의 속성명을 입력하여야 하는데, 본 프로젝트의 온톨로지는 여러 기관에서 다양한 이용자를 위해 쓰일 예정이므로 하나의 의미로 통용될 수 있는 표준언어를 구사해야 한다. 따라서 더블린코어(Dublin Core)와 스키마(Schema) 사이트를 통해 알맞은 언어를 찾았으며, 해당 단어가 생각한 정의와 부합하지 않거나 존재하지 않는 경우에는 자체적으로 제작하여 적용하였다.

온톨로지와 홈페이지에 적용할 인스턴스도 필요하기 때문에 한국어교재 사이트에서 제공하고 있는 학습자료 시리즈를 여럿 선정하여 제목, 난이도, 제공처, 저자, 기능, 영역, 구성자료 등 연결된 모든 정보를 모아 앞서 정한 기본 온톨로지 표준을 기반으로 작업하였다.

온톨로지를 토대로 홈페이지를 제작하기 위해 HTML 코딩 방법을 조사하던 도중, 코딩 없이 쉽게 홈페이지를 제작할 수 있는 사이트인 Wix 를 알게되었다. 이에 따라, 작업 방식을 변경하여 wix 의

구체적인 제작 방법을 알아보는 것으로 구축 전 조사과정을 마무리하였다.

2) 교수님 자문

본 프로젝트를 시작한 이후부터 6주차인 4월 초까지 프로젝트의 방향성에 대하여 매주 금요일마다 정경희 교수님¹께 자문을 구하였다. '누구를 위한 교재이며, 교재라고 하는 것이 적절한가, 주제가 너무 광범위하고 현실성 없지는 않은가'하는 종합적인 피드백을 토대로 주제와 예상개발자, 자료 범위 등을 한정하였다.

프로그램 초반, 인터뷰 및 설문조사 방식에 대해 노정은 교수님²께 자문을 구하였다. 이를 통해 사이트의 정체성에 대한 방향 및 세부 항목 설정에 관한 자문을 얻을 수 있었다.

온톨로지를 구축하는 과정에서 5월 25일 줌 화상통화를 통해 한상은 교수님³께 자문을 구하였다. 온톨로지의 네임스페이스 및 하위 클래스 설정, 속성명, null, 중복 입력 관련 질문을 포함해 전체적인 구조에 대한 자문을 얻어 수정 작업을 진행하였다.

2. 프로젝트 결과물 개요

2.1 프로젝트 결과물 설명

한국어 학습자료에 대해 기본적으로 제공할 정보 유형을 통합시켜 하나의 기준점이 될 온톨로지를 구축하였다. 해당 프로젝트에서 다루는 한국어 학습자료는 도서 형태인 교재와 익힘책, 어휘사전이며, 영상, 음원, 이미지 등의 모든 학습 보조자료를 포함한다. 이 온톨로지는 한국어

¹ 한성대학교(크리에이티브인문학부). 2022-1 '디지털인문정보학 캡스톤디자인' 과목 교수님.

² 한성대학교(크리에이티브인문학부). 2022-1 '한국어 평가의 원리와 실제' 과목 교수님

³ 한성대학교(크리에이티브인문학부). 2021-2 '인문콘텐츠와 정보기술' 과목 교수님.

학습자료를 소개하는 모든 공공기관과 대학 사이트 외 많은 곳에서 쓰일 수 있다. 특히 한국에 거주하는 외국인 학습자에게 가장 적극적이고 가까운 기관으로, 지역 다문화센터 및 중앙 다문화 포털 사이트인 '다누리'에서도 이를 기반으로 한국어 학습자료에 대한 정보를 소개할 수 있도록 하는 예가 될 수 있으며, 이용자의 편의를 최대한 보장하기 위해 다양한 기준에 따라 검색에 용이한 형태로 구축하였다. 온톨로지를 바탕으로 임시 제작한 홈페이지는 각 사이트의 상황에 맞추어 수정이 편리하게 이루어질 수 있도록 단순한 레이아웃과 구성만을 적용하였다.

2.2 구조 및 기능

2.2.1 표준체계 : 온톨로지 구축

단어의 의미가 우리가 말하고자하는 바를 명확하게 정의하고 있는가, 누가보아도 하나의 뜻을 내포하고 있는가를 기준으로 온톨로지 속성명은 더블린코어와 스키마를 사용해 최대한 존재하는 이름으로 사용하려 했으며, 정의가 명확하지 않거나 존재하지 않는 단어는 자체 제작하여 지정하였다.

해당 온톨로지는 검색에 용이하도록 계층 구조를 나타낸다. 학습자료가 상위 개념으로써 하위에 도서와 학습 보조자료를 둔다. 도서는 크게 교재, 익힘책, 사전을 포함하고, 난이도·저자·영역·기능·시리즈로 분류해 카테고리별로 분류했다. 난이도는 초급·중급·고급, 영역은 문법·발음·어휘·문화·기타, 기능은 종합·말하기·듣기·쓰기·읽기, 저자는 개인저자·기관저자로 구성하였다. 학습보조자료는 이미지, mp3, 동영상을 포함한다. 목차나 책소개의 경우 책을 만든 출판사에서 공식적으로 기재한 텍스트를 이용하여 보다 정확하고 통일된 정보를 제공하였다.

1) Class

클래스는 국립국어원의 분류기준과 각 사이트, 스키마 및 더블린 코어와 같은 메타데이터 표준을 참고하여 크게 8 가지로 구축하였다.

첫 번째로, 학습자료 클래스이다. 온톨로지에 입력한 명칭은 'Learning Resource'로, 스키마를 출처로 하여 작성하였다. 하위 클래스에는 학습보조자료인 'Learning Support Material'과 도서인 'Book'을 설정해두었다. 학습보조자료 클래스는 정식적인 출판물이나 도서 형태가 아닌, 개별적인 파일로 제공된 이미지나, 동영상, 음성 유형의 보조자료를 작성하기 위한 클래스이다. 도서 클래스는 도서로써 출판된 공식적 학습자료를 작성하기 위한 클래스이다. 또한, 도서의 하위 클래스에 교재인 'Textbook'과 익힘책인 'Workbook', 사전인 'Dictionary'를 설정해두었다. 도서는 학습자료와 마찬가지로 스키마를 출처로 작성하였으며, 학습보조자료와 교재, 익힘책, 사전 클래스는 제작하여 작성하였다.

두 번째로, 유형 클래스이다. 'Material Type'으로 자체 제작하여 작성하였다. 이 클래스는 학습보조자료의 유형에 대한 클래스로, 이미지, 동영상, 음성 파일 유형만을 인스턴스로써 담아내었다. 각 인스턴스의 명칭은 더블린 코어를 참고하여 image, movingImage, sound 로 작성하였다.

세 번째로, 영역 클래스이다. 'Language Domain'으로 이 역시 제작하여 작성하였다. 이 클래스는 해당 학습 자료가 다루는 영역에 대한 클래스로, 문법, 발음, 어휘, 문화, 기타, 알 수 없음을 인스턴스로 담아내었다.

네 번째로, 기능 클래스이다. 'Language Skill'로 제작하여 작성하였다. 이 클래스는 해당 학습 자료가 다루는 기능에 대한 클래스로, 종합, 말하기, 듣기, 읽기, 쓰기, 알 수 없음을 인스턴스로 담아내었다.

다섯 번째로, 난이도 클래스이다. 'Education Level'로, 스키마를 출처로 하여 작성하였다. 이 클래스는 해당 학습 자료의 난이도에 대한 클래스로, 초급, 중급, 고급을 인스턴스로 담아내어 표현하였다. 이 인스턴스 명칭은 스키마의 Education Level 서술 페이지 속 안내를 따라 beginner, intermediate, advanced 로 작성하였다.

여섯 번째로, 개인저자 클래스이다. 'Person'으로, 스키마를 출처로 하여 작성하였다. 이 클래스는 학습 자료를 제작한 저자가 개인인 경우를 표현하기 위한 클래스로, 다른 사람이 이 기준을 사용할 때 활용성과 이해도를 넓히기 위해 스키마에서 명칭을 가져와 작성하였다.

일곱 번째로, 기관저자 클래스이다. 'Organizaion'으로, 스키마를 출처로 하여 작성하였다. 이 클래스는 개인저자 클래스와 달리 기관이나 단체명으로 학습 자료를 제작한 경우를 표현하기 위한

클래스이다. 이 클래스도 개인저자 클래스와 마찬가지로, 활용성과 이해도를 위해 스키마에서 명칭을 가져와 작성하였다.

마지막으로, 시리즈 클래스이다. 'Collection'으로 스키마를 출처로 하여 작성하였다. 순서가 중요한 연속된 출판물이 아니라 같은 저자, 같은 범주 아래, 구성품으로써 함께 포함된 자료를 묶어서 표현하기 위한 클래스이다.

2) Object Property

표 1

Domain	Range	Object Property
Learning Resource	Organization, Person	is produced by
Organization, Person	Learning Resource	produces
Learning Resource	Education Level	has level
Education Level	Learning Resource	is level of
Learning Resource	Language Domain	has domain
Language Domain	Learning Resource	is domain of
Collection	Learning Resource	has part
Learning Resource	Collection	is part of
Book	Language Skill	has skill
Language Skill	Book	is skill of
Book	Learning Support Material	relates to

Learning Support Material	Book	is related to
Learning Support Material	Material Type	has type
Material Type	Learning Support Material	is type of

더블린 코어에서 가져온 is part of, has part 외에는 모두 제작하여 나타내었다.

3) Datatype Property

‘출판일’, ‘이름’은 스키마. ‘도서명’, ‘목차’, ‘책소개’, ‘자료명’은 더블린 코어를 출처로 하며, ‘구매처’는 제작하였다.

표 2

Domain (Class)	Datatype Property	Range
Book	date published	date
Book	description	string
Book	table of contents	string
Book	website to purchase	string
Learning Resource	title	string
Organization, Person	name	string

4) 제약 사항

아래의 제약 사항은 각 클래스에 owl:equivalentClass 를 통해 서술하였다.

- 저자는 적어도 하나의 학습자료를 가진다
→ produces some schema:LearningResource
- 시리즈는 적어도 하나의 학습자료를 포함한다
→ dcterms:hasPart some schema:LearningResource
- 학습자료는 하나의 난이도에만 해당해야 한다
→ hasLevel exactly 1
- 영역은 문법, 발음, 어휘, 문화, 기타, 알수없음만 존재한다
→ {grammer, pronunciation, vocabulary, culture, etcetera, none}
- 기능은 종합, 말하기, 듣기, 읽기, 쓰기, 알수없음만 존재한다
→ {comprehensiveness, speaking, listening, reading, writing, none}
- 난이도는 초급, 중급, 고급만 존재한다
→ {beginner, intermediate, advanced}
- 자료유형은 이미지, 동영상, 음성파일만 존재한다
→ {image, movingImage, sound}

5) 분류 기준

분류기준을 설정하는 과정에서 조사한 바에 따르면 예를 들어 영역이나 기능에 대해서 서점과 도서관은 관련 내용을 다루고 있지 않았으며, 출판 기관에서는 기관마다 소개 방법 및 내용이 제각각인 경우가 많이 있다. 프로젝트에서는 그 중 분류기준을 난이도, 기능, 영역, 저자, 시리즈로 선정하여 제공하려 한다.

이용자 측면에서 편의성을 제공하기 위한 사이트를 만드는 것이 본 목적이었기 때문에 이용자가 학습자료를 이용함에 있어 표면적인 내용 뿐 아니라 내부적인 요소들까지도 고려해야 이용자 중심의 정보가 제대로 전달될 수 있을 것이라 생각하였다.

주제와 같이 더 상세한 부분을 담아내기에는 개발자 입장에서 정보를 얻는 것이 현실적으로 한계가 있기 때문에 난이도·영역·기능 세 부분으로 포괄해 타협을 보았고, 난이도는 대부분 따로 나와있거나 설명 부분에서 쉽게 얻을 수 있는 정보였으므로 기본적으로 필요한 사항이라 생각되었다. 영역과 기능은 다른 것과 다르게 차별성을 두고 이용자 중심의 정보를 제공하기 위해서 실질적인 내용이 필요할 것이라 생각하였다. 각 요소들은 국내 공식적인 한국어 학습자료 개발기관인 국립국어원의 세부검색 등에서 검색 시 정보가 존재하거나 추론할 수 있는 내용들로 구성하였으며 전혀 없는 내용이면 선택지를 '알수없음'으로 하여 최대한 충분한 설명을 하려고 하였다. 시리즈는 개별 학습지 뿐 아니라 묶여있는 보조자료나 관련도서를 함께 제공받기 원하는 학습자들이 많기 때문에 선정하였다. 저자의 경우 학습자별로 선호하는 저자가 있을 수 있고, 사용하고 있는 자료와 같은 저자 시리즈를 사용하고 싶어하여 분류기준으로 삼았고 알아보기 쉽도록 개인저자와 기관저자로 구분하였다.

온톨로지를 다른 기관에서 이용하고자 할 때 오해없이 우리의 의도대로 정보를 제공하기 위해서는 누구나 이해할 수 있는 정확한 표준을 사용해야 했기에 기존에 있는 표준언어인 경우 출처를 같이 작성했지만 없는 경우 자체적으로 제작하였다.

6) 실제 온톨로지 구축 모습

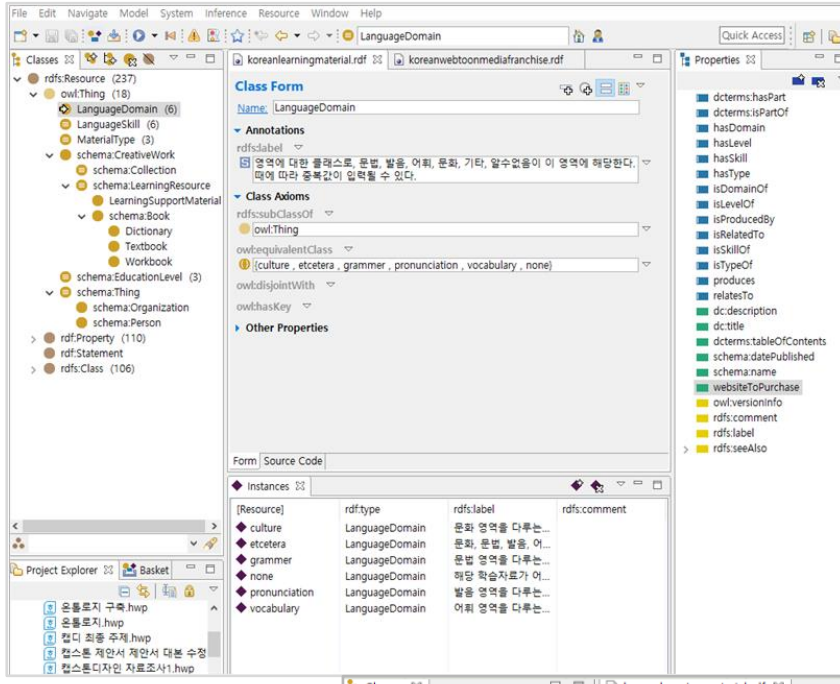


그림 1

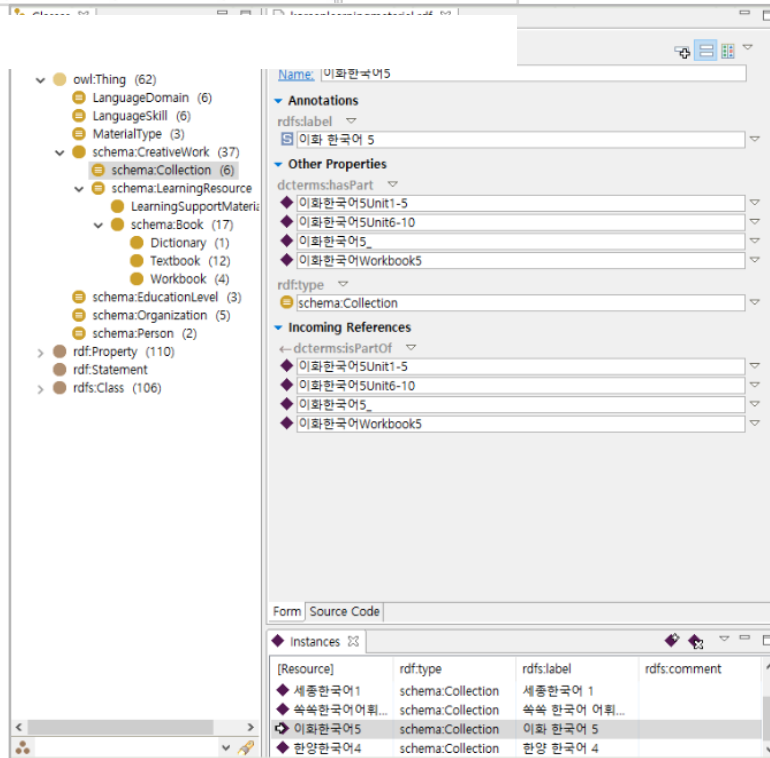


그림 2

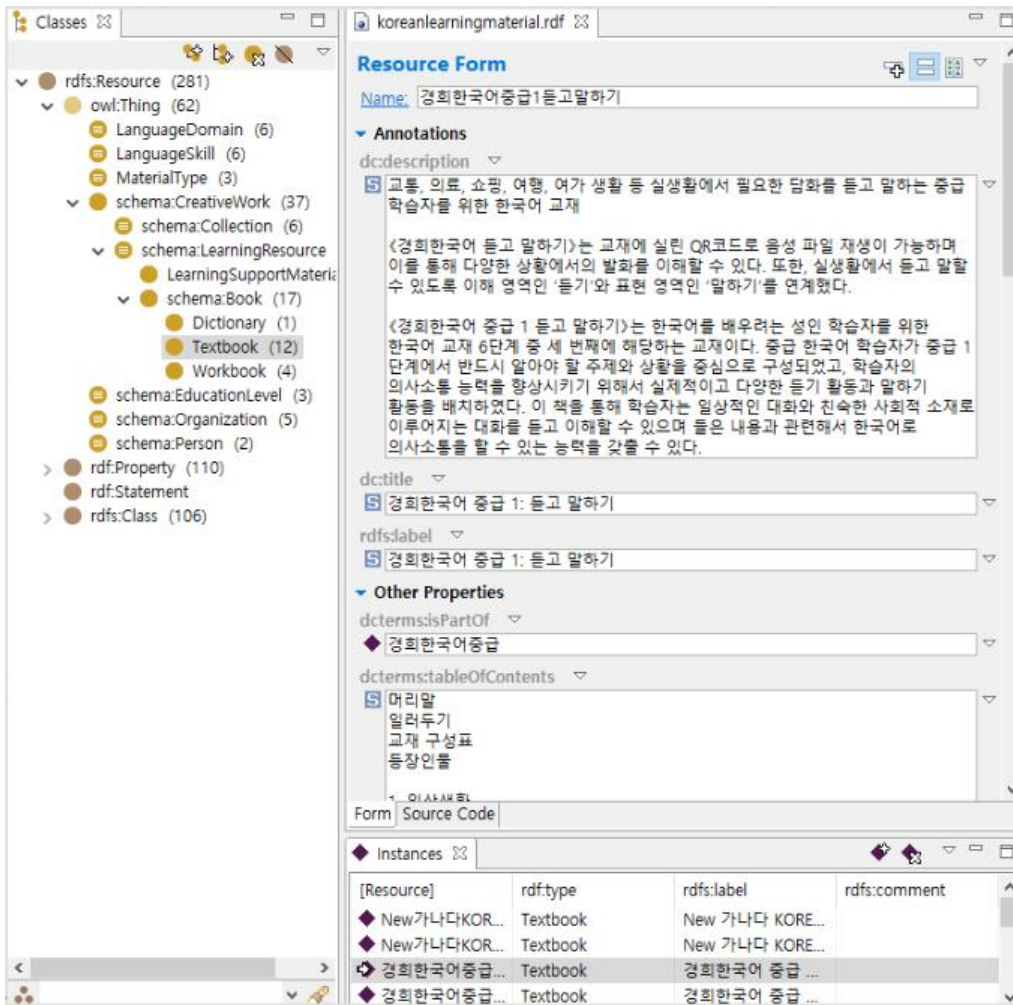


그림 3

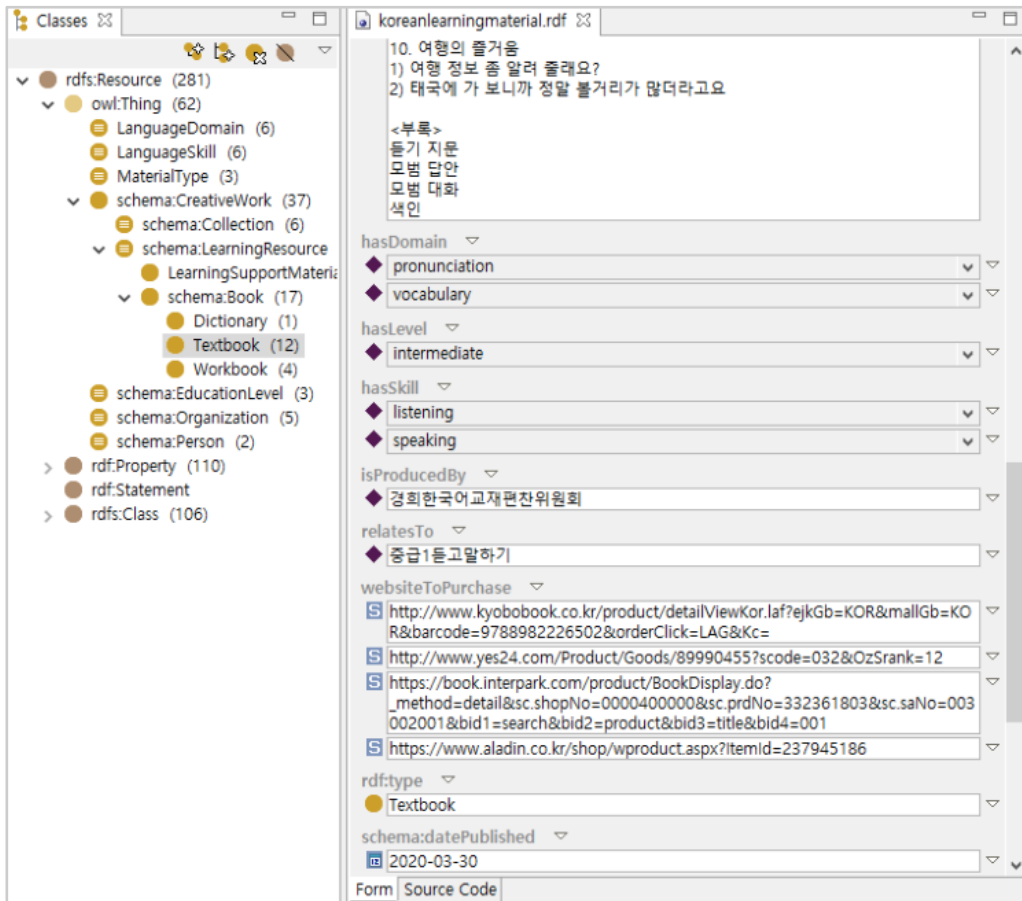


그림 4

2.2.2 적용예시 : 홈페이지 구축



그림 5

사진의 로고는 해당 사이트가 한국어 학습을 원하는 외국인을 위해 한국어 학습자료를 소개하는

기준을 제시함에 따라 한교도서관에서 관련 도서들을 소개 대상으로 다루고 있음을 표현하는 의미에서 펼쳐진 책 모양을 포함하였다. 또한 '한교도서관'의 약자로 'ㅎ', 'ㄱ', 'ㄷ'을 펼친 책 위에 배치하였다. 영어를 포함하지 않고 한글의 자음 조합을 사용하여, 한국어를 다루는 사이트로서의 정체성을 보여줄 수 있도록 표현하였다.

홈페이지의 구조는 크게 '메인화면/ 도서-교재, 익힘책, 어휘사전/ 학습 보조자료-이미지, 동영상, 음성/ 도서 하위페이지별 카테고리- 저자별, 기능별, 영역별, 난이도별, 시리즈별/ 자료 상세페이지'로 구분해 볼 수 있다.

1) Home

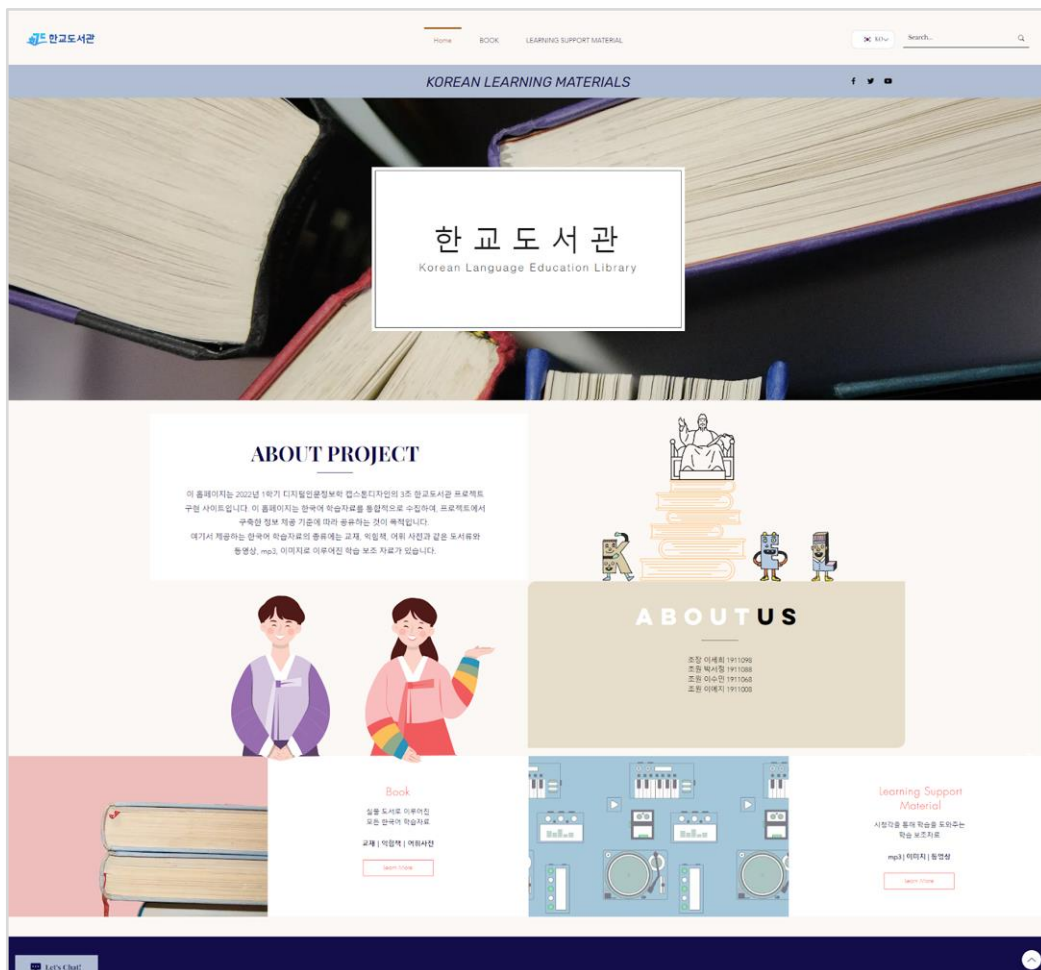


그림 6

<Home> 페이지로, 홈페이지의 첫 화면이다. 한교도서관 로고와 메뉴바, 사이트 번역기능, 특정 자료 검색을 위한 검색창을 상단에 배치하고, 화면을 스크롤해도 보일 수 있도록 모든 페이지에

고정시켜두었다. 메뉴바는 Home 과 Book(교재), book 의 하위로 TextBook(교재)·Workbook(익힘책)·Dictionary(어휘사전)이 있고, LearningSupportMaterial(학습보조자료)과 하위 Sound(음성파일)·Image(이미지)·MovingImage(동영상)으로 구성하였다. 교재 각 하위마다 있는 카테고리 페이지도 메뉴바로 구성하려 하였지만 워스에서 가능한 기능이 없어 하지 못했다. 사이트 번역은 대표적으로 한국어·영어·일본어·러시아어·중국어로 구성하였으며, 필요하다면 다른 국가언어를 추가할 수 있다.

그 아래로 도서 이미지를 스트립에 추가하여 한국어와 관련된 홈페이지의 전체적인 느낌을 살려주었다.

메인 섹션에는 세종대왕, Korean Education Library 의 이니셜, 한복입은 사람 아이콘과 함께 홈페이지에 대한 소개와 목적, 팀원 정보를 보여준다. <Book>과 <Learning Support Material> 페이지로 바로 이동할 수 있는 버튼도 추가하였다.

가장 하단에는 챗봇 기능과 위로가기 기능을 모든 페이지에 구현하여 정보 탐색에 있어 편리함을 더하였다.

2) Book/ Learning Support Material

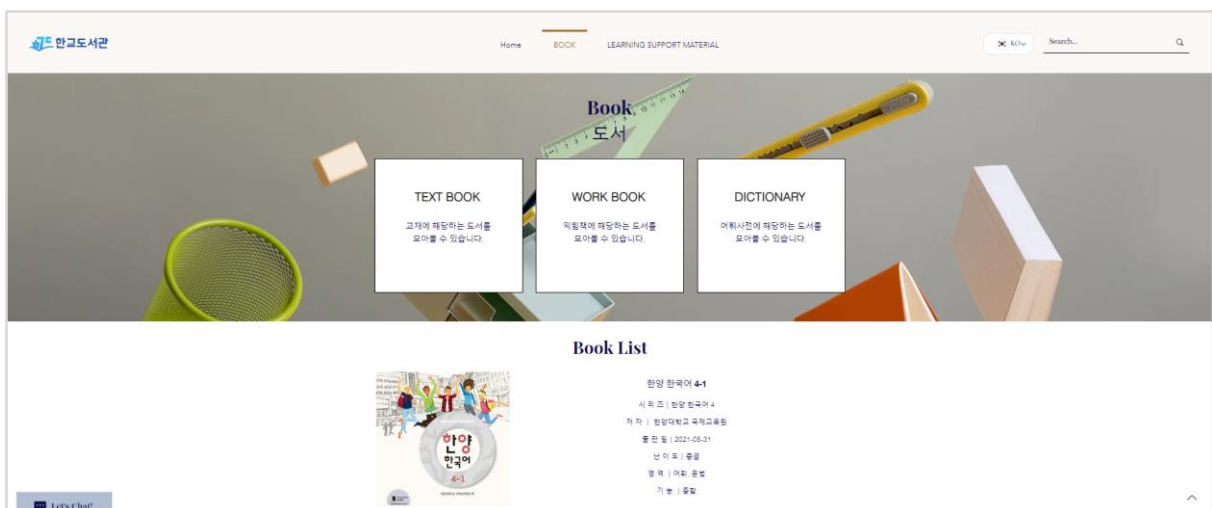


그림 7

<Book> 페이지는 실물 도서만을 포함하고 있으며 <Text Book>, <Work Book>, <Dictionary> 이동 버튼 상자 아래 도서 목록을 나열하였다.

도서 목록의 이미지 혹은 텍스트를 클릭하면 해당 도서의 상세페이지로 이동할 수 있도록 링크하였으며, 이미지 옆에는 제목, 시리즈, 저자, 출판일, 난이도, 영역, 기능 등 간단한 도서 정보를 볼 수 있게 나타내었다.

<Learning Support Material> 페이지도 이와 동일하다.

3) TextBook/ Workbook/ Dictionary, Sound/ Image/ MovingImage

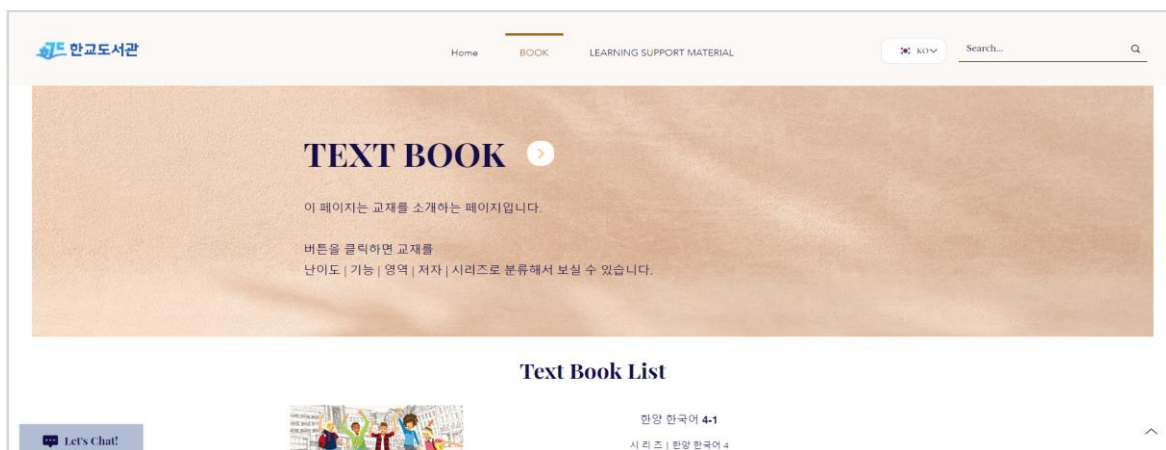


그림 8

다음 사진은 도서 하위의 <Text Book> 페이지로, 제목 옆에 카테고리 페이지로의 이동 버튼을 구현해 놓았다. 아래는 교재목록을 나열하였다. work book, dictionary 페이지도 이와 동일하다.

학습보조자료 하위의 Sound, Image, Moving Image 페이지도 비슷하지만, 한 가지 차이는 카테고리 이동 버튼이 없다는 것이다. 우리가 구축한 홈페이지에는 주로 교재의 시리즈와 연관된 보조자료들을 적용하였지만, 학습보조자료는 대체로 도서와 달리 도서에 포함되어 있거나 개별적으로 존재하는 자료, 정보가 얼마 없는 자료 등 서점이나 사이트들에서 나타내는 유형·형태·제공 방식이 무척 다양하다. 따라서 이를 도서의 분류기준과 같이 나눈다면 오히려 이용자의 자료 선택 과정에서 혼란을 줄 수 있다고 판단하여 제공하지 않았다.

도서 목록이 이미지, 정보와 함께 아래로 나열되어 있는 다른 페이지와는 다르게, <Image> 페이지는 시각적인 요소가 부각되어야 한다고 생각하여 갤러리 형식의 레이아웃을 사용해 이미지, 제목만 나타내어 목록을 구성하였다.

4) Category

도서의 하위페이지마다 카테고리 페이지를 생성하여 교재·익힘책·어휘사전의 도서를 또 다시 저자, 난이도, 영역, 기능, 시리즈별로 분류해서 볼 수 있다. 이는 사전 조사 과정 중 여타 한국어 교재 사이트들에서 가장 많이 소개가 된 것 또는 사람들이 한국어 교재를 찾을 때 가장 중요하게 보는 것이라고 생각되는 기준을 종합하여 선정한 분류기준이다. 각각의 이동 버튼을 만드는 대신 하나의 페이지에 모두 나타내어 스크롤을 내리거나 우측의 앵커를 클릭하면 원하는 카테고리로 화면이 움직인다.

각 카테고리는 앞서 온톨로지에서 언급한 요소들을 포함하고 있으며, 요소를 클릭하면 해당 도서 목록을 보여준다.



그림 9

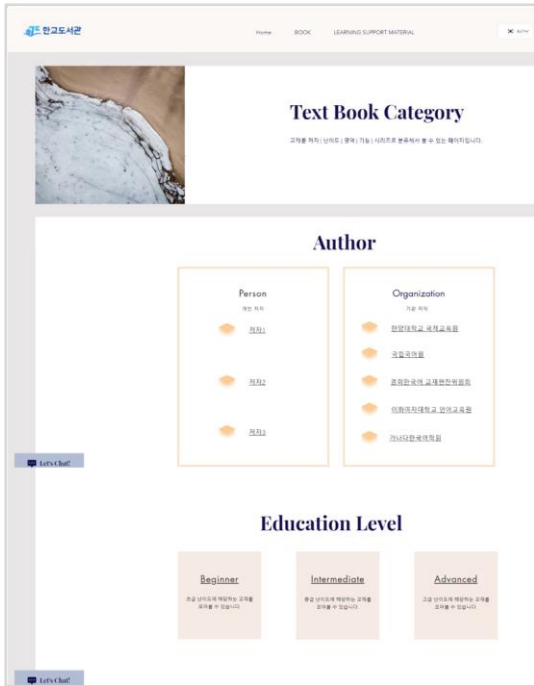


그림 10

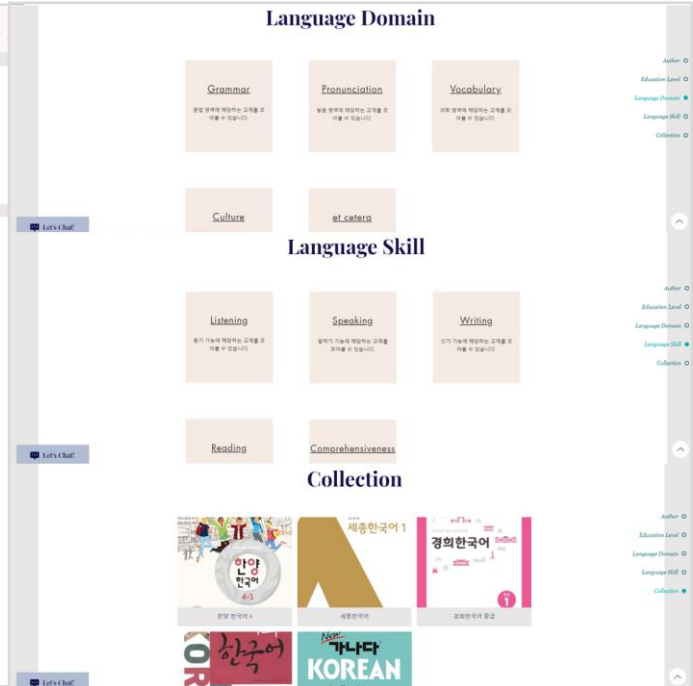


그림 11

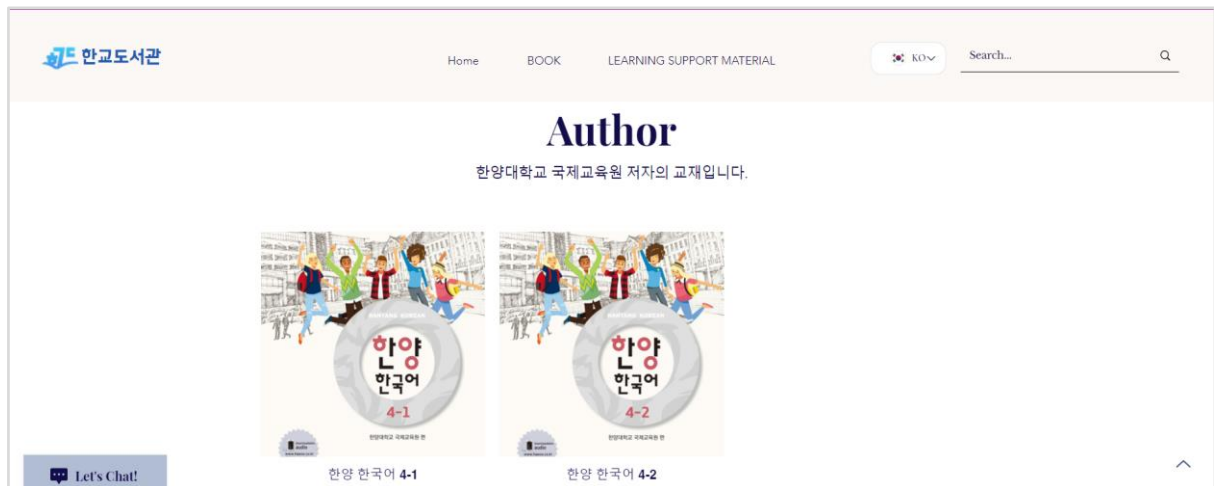


그림 12

5) 상세 페이지

도서 상세 페이지는 다음과 같다. 목록에서와 같이 자료의 이미지, 제목, 시리즈, 저자, 출판일, 난이도, 영역, 기능을 보여주며, 추가로 구매처 링크를 제공하여 구매하고자 하는 도서라면 미리보기,

가격 등 더 상세한 정보를 찾아볼 수 있게 돕는다. 구매처는 도서자료에만 제공될 뿐, 학습보조자료는 판매하는 상품이 아니므로 적용하지 않았다. 학습보조자료는 정확한 출판일이나 개정일 등을 알 수 없고 기능 부분을 판단할 수 없어 출판일과 기능 또한 제외하는 대신, 해당 자료의 유형을 작성하였다.

그 아래 도서 상세 페이지의 기본적인 구성은 책소개, 목차와 함께 공부하면 좋을 구성 자료들, 해당 도서를 포함하는 시리즈 도서 및 학습보조자료를 보여주고 있다. 책 소개는 도서를 출판한 해당 대학이나 기관의 것을 사용하였다. 동영상은 책소개와 목차 대신 유튜브 영상을 삽입하였고, 유튜브가 아닌 다른 곳의 영상 자료일 경우 외부 링크를 통해 보여주도록 한다. mp3의 경우, 해당 페이지에서 바로 음성 파일을 들을 수 있도록 구축하였다.

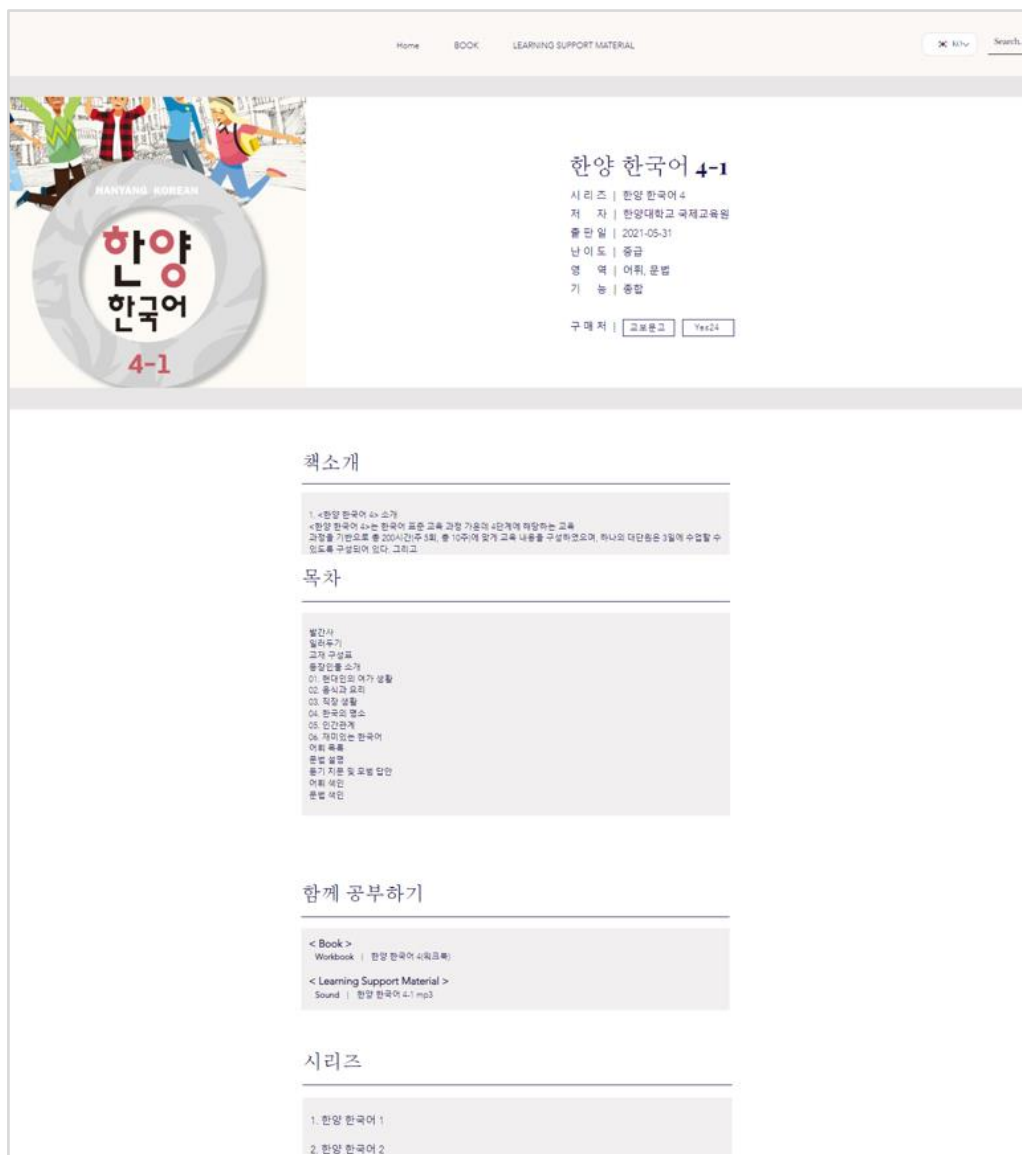


그림 13

가능하다.

본 프로젝트에서 구축한 홈페이지는 온톨로지 속 기준을 토대로 사이트에 실제 적용시켰을 때 어떤 방식과 모습으로 제공될 수 있는지 참고할 수 있다. 또한, 해당 프로젝트에서는 제작하지 않았지만 모바일 버전의 홈페이지도 존재하기 때문에 사용 기기에 상관없이 알맞는 홈페이지를 제공할 수 있다.

2) 단점

온톨로지는 팀원 모두 해당 분야의 전문가가 아니므로 지식적인 측면에서 부족함이 있고, 실질적인 업무 방식이나 구현 방식에 대해 고려하지 못했기 때문에 국립국어원, 세종학당, 각 대학교의 한국어 교육센터, 그리고 다문화센터 사이트 등 실제 현장에서 그대로 적용되기엔 기술적인 한계점이 존재한다. 다양한 기관에서 사용할 수 있도록 용어를 통일하고 기준 및 관계·구조 등을 어떻게 나타내고 있는지 보여줄 수 있도록 아이디어만 실체화하였다. 그리고 실제로는 더욱 다양한 유형과 범위의 자료를 모두 다루지 못하였다.

홈페이지도 마찬가지로 온톨로지를 활용해 만든 시각자료로서 보여질 뿐, 실제 사용을 위해 코딩한 것이 아니라 간단하게 디자인을 하여 실제와 비슷하게 기본적인 구성만을 갖추고 있다. 또한 페이지 좌측에 원하는 페이지로 바로 이동 가능하도록 선택창을 두어 이용자의 편의성을 높이려 하였다. 이 같은 방식은 다문화 센터 '다누리' 및 부산 다문화교육 지원센터 홈페이지 등 많은 기관에서 사용하여 참고한 것이지만, 그러한 선택창을 만들 수 있는 기능이 없었으며 홈페이지 자체적인 뒤로가기 기능조차 지원하지 않아 페이지 이동에 있어 불편함이 존재한다. 또한, 도메인을 확보하지 못하여 프로젝트와 관련없는 형태의 링크가 작성되었다.

이러한 단점을 보완하여 차후 프로젝트에서는 학습자료의 유형을 확장하여 구축하고 도메인 구입 및 정적 페이지 등의 홈페이지 구축 지식을 공부하여 보다 전문적인 홈페이지의 모습으로 개선하는 것이 필요하다.

2.5 기대효과

표준화된 하나의 온톨로지를 통해 한국어 학습자료에 대한 정보를 사이트마다 일정하게 전달할 수 있게 되면, 해당 기관의 홈페이지를 이용해 한국어 학습자료에 대한 정보를 얻어야 하는 이용자의 입장에서는 분산된 정보환경에서 오는 혼란을 제거함으로써 어려움이나 혼란 없이 해당 사이트에서 제공하는 정보를 쉽게 받아들일 수 있고 빠른 검색이 가능할 것이다.

더불어 지역 다문화 센터에서도 한교도서관의 온톨로지를 이용해 관련 정보를 제공하게 된다면, 한국어 교실과 같은 지역 프로그램에 참여하는 한국어 교사 및 예비 교원과 봉사활동 참여자, 홀로 한국어를 독학하고자 하거나 한국어 학습 과정에서 더 많은 자료를 얻고 싶은 외국인 등 다양한 상황에서 한국어 자료를 필요로 하는 학습자들에게 여러 사이트에 흩어져 있는 정보를 직접 수집하고 비교해야 하는 불편함이 없도록 도움을 줄 수 있을 것이다.

또한 관리자 입장에서도 어떤 정보로 제공해야 할지에 대한 고민을 감소시킬 것으로 기대되며, 효율적으로 최대한의 자료를 정확하게 보여줄 수 있게 되고, 데이터를 기존의 시스템보다 손쉽게 추가·삭제하는 것이 가능해진다.

2.6 관련 기술 및 개발방법

2.6.1 관련 기술

1) 온톨로지

온톨로지의 대표적인 정의로는 토마스 그루버(Thomas A. Gruber)의 '공유되는 개념화의 형식적이고 명시적인 명세'가 있다(1993). 즉, 온톨로지란 사물 등 세상에 존재하는 것에 대하여 일반적으로 공유되고 합의된 개념과 관계를 컴퓨터가 읽고 처리할 수 있는 형태로 표현한 것을 말한다. 이는 문헌정보학을 비롯해 시맨틱 웹, 자연어 처리 등 다양한 분야에서 활용되고 있다.

온톨로지는 서로 다른 시스템 간 기계적 처리가 가능하도록 공통 개념을 제공할 수 있도록 하여 데이터 공유나 재사용, 재활용이 가능하도록 한다. 정보검색의 측면에서 바라본다면 의미 기반의 검색이 이루어짐으로써 검색자의 의도를 반영하는 지능적 정보검색 서비스를 제공할 수 있도록

지원한다.

온톨로지는 클래스(class), 속성(property), 인스턴스(instance), 관계(relation)로 구성된다. 클래스는 사물 또는 개념에 해당하는 것으로 온톨로지의 핵심요소이며, 인스턴스는 클래스에 해당하는 실질적 사례를 말한다. 속성은 클래스나 인스턴스가 가진 특성, 성질 등을 의미하며, 관계는 클래스·인스턴스 간의 관계를 나타낸다.

2) 데이터베이스

여러 사람에 의해 공유되어 사용될 목적으로 통합하여 관리되는 데이터 집합을 말한다. 논리적으로 연관된 하나 이상의 자료의 모음으로, 그 내용을 고도로 구조화함으로써 검색과 갱신의 효율화를 꾀한 것이다. 즉, 몇 개의 자료 파일을 조직적으로 통합하여 자료 항목의 중복을 없애고 자료를 구조화하여 저장해 놓은 '자료의 집합체'라고 할 수 있다.

데이터베이스가 갖는 몇 가지 특성을 살펴보자면, 첫 번째로는 똑같은 자료를 중복하여 저장하지 않는 통합된 자료이다. 두 번째로는 컴퓨터가 접근해 처리할 수 있는 저장 장치에 수록된 자료이다. 세 번째로는 어떤 조직의 기능을 수행하는 데 있어 없어서는 안되며, 존재 목적이 뚜렷하고 유용성 있는 운영자료이기 때문에 임시로 모아놓은 데이터나 단순한 입출력 자료가 아니라는 점이다. 마지막으로, 한 조직에서 가지는 데이터베이스는 그 조직 내의 모든 사람들이 소유·이용·유지하는 공동 자료로서 각 사용자는 같은 데이터라 할지라도 각자의 응용 목적에 따라 다르게 사용할 수 있다.

2.6.2 개발 도구

1) TopBraid Composer

TopBraid Composer 는 미국의 시맨틱 웹 솔루션 업체 topquadrant 에서 제작한 IDE(효율적으로 소프트웨어를 개발하기 위한 통합 개발환경 소프트웨어 어플리케이션 인터페이스⁴)로, 온톨로지 모델링 도구이자, SPARQL 쿼리 에디터, 데이터 통합 콘솔 및 시맨틱 어플리케이션 구축에 쓰이는

⁴ [네이버 지식백과] IDE [Integrated Development Environment] (두산백과 두피디아, 두산백과)

도구이다. W3C 표준에 준거하여 온톨로지 및 링크 데이터의 구성 개발, 관리 및 테스트를 지원한다.

이 IDE 는 2021 년 2 학기, 윤소영 교수님의 '시맨틱데이터 구축'에서 알게 된 도구로써, 이 프로젝트만의 목적에 맞게 더 풍부한 내용을 담고 데이터 관리 및 검색에 용이한 온톨로지를 구축하고자 한다.

2) Wix

자신의 홈페이지를 무료로 제작하고 링크까지 만드는 이스라엘 사이트이다. 본인의 성향대로 홈페이지를 제작할 수 있으며 다양한 무료 템플릿을 제공하고 있어, 홈페이지 제작에 미숙한 이용자도 쉽게 접근할 수 있다. Wix ADI 인공지능으로 홈페이지 자동 디자인이 가능하고 Wix 에디터를 사용하여 간단하게 제작할 수 있다. 현재, 웹디자인 전문가를 위한 고급 솔루션인 Velo 도 지원하고 있다.

3. 프로젝트 추진 체계 및 일정

1) 역할분담

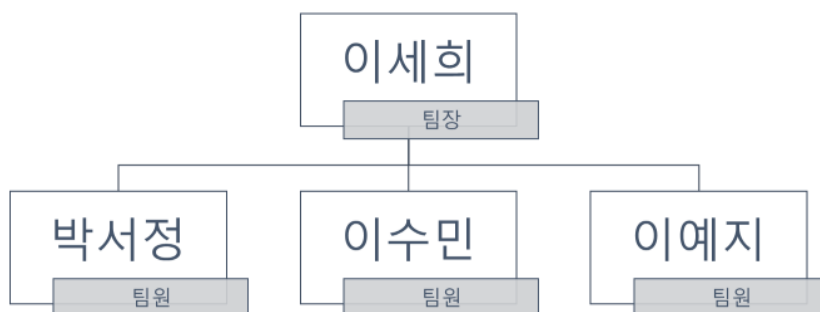


그림 16

표 3

팀원	역할	담당
이세희	팀장	온톨로지 구축 및 인스턴스 입력

박서정	팀원	홈페이지 로고 디자인 인스턴스 수집·홈페이지 입력
이수민	팀원	인스턴스 수집
이예지	팀원	홈페이지 구축

1) 수행일정표

표 4

구분		담당자	3 월				4 월				5 월				6 월			
주제선정		팀																
사이트 및 교재 현황 조사		팀																
운틀로지	관련 기술 학습	팀																
	패킷(속성) 조사	팀																
	표준언어 조사	박서정, 이수민, 이예지																
	인스턴스 수집	박서정, 이수민																
	구축	이세희																
	인스턴스 입력																	
홈페이지	관련 기술 학습	팀																

	로고 디자인	박서정																	
	구축	이예지																	
	인스턴스 입력	박서정																	
주간회의록 작성		이예지																	
제안서 작성		팀																	
보고서	작성	이수민, 이예지																	
	발표 자료	박서정, 이세희																	
한성대 위키 작성		이수민																	

4. 참고문헌 및 참고한 웹 사이트

1) 참고문헌

data base, IT 용어사전,

<https://terms.naver.com/entry.naver?docId=1082446&cid=40942&categoryId=32840>

'Ontology', IT 용어사전,

<https://terms.naver.com/entry.naver?docId=863084&cid=42346&categoryId=42346> (2022.03.31)

'Ontology', 지형 공간정보체계 용어사전,

<https://terms.naver.com/entry.naver?docId=3480314&cid=58439&categoryId=58439> (2022.03.31)

Topbraid Composer 사이트, <https://www.topquadrant.com/#>, 2022.04.04

‘다문화학생 한글교육·진학 지원... 동등한 출발선 보장’, 뉴시스, 2022.02.04,
https://newsis.com/view/?id=NISX20220204_0001746659&cID=10201&pID=10200#

백운용 기자, ‘대구시, 대구거주 외국인주민 대상 한국어 무료교육 실시’, 국제뉴스, 2022.01.28.,
<https://www.gukjenews.com/news/articleView.html?idxno=2398858>

윤소영, 2021, ‘4-1 지식표현-온톨로지’, 디지털인문정보학트랙, 한성대학교

이소희 기자, ‘수원시, 다문화가족 자녀 한국어 학습 방문 교육 지원...대상자 모집’, 민중의 소리,
 2022.02.03, <https://vop.co.kr/A00001608411.html>

이희용 기자, ‘외국인 유학생 유치 위해 해외 한국어 공교육 강화해야’, 연합뉴스, 2017.07.07,
<https://www.yna.co.kr/view/AKR20170707083200371?input=1195m>

이희용 기자, “해외 한국어교육 문제없나...유학생 75% ‘사교육 독학 의존’”, 연합뉴스, 2017.07.07,
<https://www.yna.co.kr/view/AKR20170707101600371?input=1195m>

임성준 기자, “세종시교육청, 다문화 학생 ‘한국어 맞춤형 캠프’ 운영”, 매일경제 TV,
 2022.01.17, http://mbnmoney.mbn.co.kr/news/view?news_no=MM1004533486

임성준 기자, ‘제주교육청, 다문화 가정 등에 맞춤형 통합 지원 강화’, 세계일보, 2022.01.27,
<http://www.segye.com/newsView/20220126506528?OutUrl=naver>

임영은 기자, ‘소외됐던 외국인 주민 체계적 지원’, 충북일보, 2022.03.16,
<https://www.inews365.com/news/article.html?no=706617>

2) 참고 사이트

DCMI, <https://www.dublincore.org/>

Schema.org, <https://schema.org/docs/schemas.html>

SD 에듀(주)시대고시기획, “<쏙쏙 한국어 어휘왕 TOPIK 1 초급>MP3 파일입니다.”,
 2022 년 06 월 13 일접속,
https://www.sdedu.co.kr/bbs/board.php?bo_table=data_mp3&wr_id=426&svs=book&page=3

SD 에듀(주)시대고시기획, “[향균안심도서]쏙쏙 한국어 어휘왕 TOPIK 1 초급 단어사전”,
 2022 년 06 월 13 일접속,
https://www.sdedu.co.kr/book/item.php?it_id=1621910056&cat_id=

<http://www.ewhapress.com/ewhapress/149/subview.do?enc=Zm5jdDF8QEB8JTJGYm9vayUyRmV3aGFwcmVzcyUyRjQ2NDIIMkZ2aWw3LmRvJTNGc2l0ZUIkJTNEZXdoYXBvZXNzJTl2bWVudVNIcSUzRDE0OSUyNnBhZ2UIM0QxJTl2ZmluZFR5cGUIM0R0aXRzZSUyNmZpbmRXb3JkJTNEJUVDJTIEJUJUVEJTk5JTk0KyVFRCU5NSU5QyVFQSVcNSVBRCVFQyU5NiVCNCt3b3JrYm9vays1JTl2>

한글파크출판, "New 가나다 KOREAN 고급 1", 2022 년 06 월 13 일접속,

https://www.sisabooks.com/hangeul/shop/view?book_num=102901

한글파크출판, "New 가나다 KOREAN 고급 2", 2022 년 06 월 13 일접속,

https://www.sisabooks.com/hangeul/shop/view?book_num=102973

한글파크출판, "New 가나다 KOREAN for foreigners 워크북 고급 1", 2022 년 06 월 13 일접속,

https://www.sisabooks.com/hangeul/shop/view?book_num=102903

한글파크출판, "New 가나다 KOREAN for foreigners 워크북 고급 2", 2022 년 06 월 13 일접속,

https://www.sisabooks.com/hangeul/shop/view?book_num=102986