

ChatGPT가 자동 생성한 더블린 코어 메타데이터의 품질 평가: 국내 도서를 대상으로

Quality Evaluation of Automatically Generated Metadata Using ChatGPT: Focusing on Dublin Core for Korean Monographs

김선욱 (SeonWook Kim)*

이혜경 (HyeKyung Lee)**

이용구 (Yong-Gu Lee)***

초 록

이 연구의 목적은 ChatGPT가 도서의 표지, 표제지, 판권기 데이터를 활용하여 생성한 더블린코어의 품질 평가를 통하여 ChatGPT의 메타데이터의 생성 능력과 그 가능성을 확인하는 데 있다. 이를 위하여 90건의 도서의 표지, 표제지와 판권기 데이터를 수집하여 ChatGPT에 입력하고 더블린 코어를 생성하게 하였으며, 산출물에 대해 완전성과 정확성 척도로 성능을 파악하였다. 그 결과, 전체 데이터에 있어 완전성은 0.87, 정확성은 0.71로 준수한 수준이었다. 요소별로 성능을 보면 Title, Creator, Publisher, Date, Identifier, Right, Language 요소가 다른 요소에 비해 상대적으로 높은 성능을 보였다. Subject와 Description 요소는 완전성과 정확성에 대해 다소 낮은 성능을 보였으나, 이들 요소에서 ChatGPT의 장점으로 알려진 생성 능력을 확인할 수 있었다. 한편, DDC 주류인 사회과학과 기술과학 분야에서 Contributor 요소의 정확성이 다소 낮았는데, 이는 ChatGPT의 책임표시사항 추출 오류 및 데이터 자체에서 메타데이터 요소용 서지 기술 내용의 누락, ChatGPT가 지닌 영어 위주의 학습데이터 구성 등에 따른 것으로 판단하였다.

ABSTRACT

The purpose of this study is to evaluate the Dublin Core metadata generated by ChatGPT using book covers, title pages, and colophons from a collection of books. To achieve this, we collected book covers, title pages, and colophons from 90 books and inputted them into ChatGPT to generate Dublin Core metadata. The performance was evaluated in terms of completeness and accuracy. The overall results showed a satisfactory level of completeness at 0.87 and accuracy at 0.71. Among the individual elements, Title, Creator, Publisher, Date, Identifier, Rights, and Language exhibited higher performance. Subject and Description elements showed relatively lower performance in terms of completeness and accuracy, but it confirmed the generation capability known as the inherent strength of ChatGPT. On the other hand, books in the sections of social sciences and technology of DDC showed slightly lower accuracy in the Contributor element. This was attributed to ChatGPT's attribution extraction errors, omissions in the original bibliographic description contents for metadata, and the language composition of the training data used by ChatGPT.

키워드: 메타데이터, 더블린 코어, 자동 추출, 자동 생성, 챗GPT

metadata, Dublin Core, automatic extraction, automatic generation, ChatGPT

* 경북대학교 사회과학대학 문헌정보학과 강사(seonwook.kim@knu.ac.kr) (제1저자)

** 경북대학교 사회과학대학 문헌정보학과 강사(keillee@knu.ac.kr) (공동저자)

*** 경북대학교 사회과학대학 문헌정보학과 부교수(yglee@knu.ac.kr) (교신저자)

■ 논문접수일자: 2023년 5월 15일 ■ 최초심사일자: 2023년 6월 3일 ■ 게재확정일자: 2023년 6월 8일

■ 정보관리학회지, 40(2), 183-209, 2023. <http://dx.doi.org/10.3743/KOSIM.2023.40.2.183>

※ Copyright © 2023 Korean Society for Information Management

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

1. 서론

20세기 컴퓨터의 발명 이래로 현대 사회는 거대한 발전을 이룩하면서, 전 세계는 인공지능이라는 신기술을 앞다투어 다양한 산업에 적용하고자 부단한 노력을 기울이고 있다.

2022년 11월, OpenAI사는 이들이 개발한 GPT 모델에 기반하여 사용자와 대화 형태로 인공지능기술을 사용할 수 있도록 하는 ChatGPT 서비스를 전 세계에 공개하였다. 이로써 비단 전문가나 연구자 한정만이 아닌 일반의 대중들도 별도의 코딩이나 AI 지식이 부족하더라도 자연어처리, 딥러닝 등으로 언급되는 AI (artificial intelligence) 기술을 수월하게 경험할 수 있는 계기가 되었다. 다시 말하면 복잡하게만 여겨지던 AI의 기술이 대중들의 실생활에도 자연스럽게 스며들어 최신기술을 누구나 이용할 수 있는 환경이 구축될 수 있으므로 현재보다 급격한 속도로 산업의 변화를 가져올 것으로 보인다.

언제나 새로운 서비스와 기술의 등장, 그리고 이에 대한 대응은 연구의 방향이나 시야를 넓혀주는 원동력으로 작용하고 더 나아가 연구의 패러다임을 전환하게 되는데, ChatGPT의 발표 이래로 각 학계에서는 이 기술이 연구나 사회에 미치는 영향, 미래의 방향성, 접목 및 활용성과 같은 연구 성과를 신속하게 발표하고 있다.

이에 발맞추어 전 세계 문헌정보학계에서도 ChatGPT 관련 연구를 수행하고 있을 것으로 예상하나, 지금까지 국내 문헌정보학계에서는 뚜렷하게 관련 연구를 찾아보기 힘들다.

본 연구는 ChatGPT의 역할을 '사서'라고 지

정하고 메타데이터 생성에 필요한 최소한의 정보를 입력했을 때 더블린 코어 메타데이터를 생성할 수 있는가? 그리고 생성할 수 있다면, 그 품질은 어느 정도인가를 분석하고자 하였다. 구체적으로 ChatGPT의 개괄과 더불어 서지정보 데이터를 기반해 ChatGPT를 활용한 더블린 코어 작성 실험을 수행하여 그 품질을 평가함으로써, ChatGPT에 대한 문헌정보학계와 도서관계의 적용 가능성을 파악해보고자 한다.

2. 선행연구

ChatGPT는 OpenAI의 대규모 언어모델 GPT3 시리즈와 GPT4 시리즈를 기반으로 하는 챗봇 서비스이다. AI의 발전은 다양한 영역에서의 활용을 위한 다각적이고 심도 있는 연구 성과를 도출하고 있는데, 특히 최근 등장한 ChatGPT의 파급력은 놀라울 정도이다. 이 여파는 대중뿐만 아니라, 전 세계 연구자들에게도 단연코 인공지능의 화두로 그 관심과 이목을 집중시키고 있다.

이에 발맞춰 최근 국내에서도 ChatGPT 관련 연구가 등장하기 시작했으나, ChatGPT의 역할이나 관련 기사, 활용 사례 등을 수집하고 각 학계에 알려, 연구범위의 확장을 넓혀주는 초석 역할로서의 연구가 주로 수행되고 있음을 확인하였다.

이윤희, 김창식, 안현철(2023)은 ChatGPT의 개괄적인 현황을 파악하고, 적절한 사용 방향을 제시하기 위하여 관련 연구에 대한 질의 응답, 기계 번역, 프로그래밍 작성, 이미지 및 음악 생성 등에서의 ChatGPT의 활용 사례를

확인하였고, 연구자와 각 집단 및 기관은 지속해 ChatGPT와 관련한 다양한 경험을 누적할 필요가 있으며, 이로써 흥미로운 연구 성과나 업무 효과를 도출될 것이라고 하였다.

송학준, 송형용, 이지은(2023)은 관광산업에서 ChatGPT가 기여 혹은 접목할 수 있는 관점을 도출하고 추후 연구영역의 식별을 위하여, 관련 산업의 연관성 측면에서 장단점을 파악하였다. 더불어 기존의 챗봇 시스템과 달리 ChatGPT는 대화의 맥락에 따라 이용자에게 맞춤형 응답을 제공할 수 있으므로 이용자와의 상호작용 방식의 혁신을 기대하고 나아가 향상된 서비스로의 발전 가능성을 제시하였다.

정한민, 박정훈(2023)은 연구의 문헌 작성 단계에서의 ChatGPT의 효율성을 확인하고자, ‘정보서비스의 개념, 그리고 주요 기능과 알고리즘’이라는 주제로 문헌 구조설계, 질문 작성, 단락 생성, 영한 번역, 교정 부문에서 등장 단어의 수, 정보의 일치성 등을 파악해 ChatGPT의 성능을 확인하였다. 그 결과 연구보고서에서의 연구 주제 선택, R&D 계획 수립에는 대체로 활용이 가능할 것으로 보이나, 수치에 대한 오류, 참고문헌 등의 부재에 따른 정보원에 대한 신뢰성 결여 등의 한계가 밝혀짐에 따라 데이터 검증에 보다 주의를 신경 써야 함을 피력하였다.

노대원(2023)은 ChatGPT를 연구과정 참조, 논문 내용 생성 및 확장, 참고 기사 요약, 문학 텍스트 작성 및 해석, 영문 초록 교정, 소설 인용문의 띄어쓰기 교정 등에 적용하여 문학의 전반적인 영역에서의 영향력을 파악하였는데, ChatGPT는 기존 문학의 소재나 주제뿐만 아니라 유명 작가의 집필 형식을 참조하고,

나아가 문학 텍스트를 창작하는 등의 결과를 보여 그 잠재력을 확인하였다. 따라서 인문학에서도 이러한 AI의 기술에서 벗어나기는 힘들 것으로 보여 인간이 주도하는 디지털 인문학 등의 융합적 발전이 시도되어야 함을 주장하였다.

위와 같이 국내 다양한 연구영역에서 ChatGPT 활용을 위한 탐색적 연구가 속속 등장하고 있음에도 불구하고, 아직 국내 문헌정보학 분야에서는 관련 연구에서는 확인할 수 없었다.

한편 해외의 경우, 2023년에 들어서 본격적으로 문헌정보학 분야의 ChatGPT 관련 연구를 살펴볼 수 있었는데, ChatGPT의 특성 탐색 및 도서관에서의 개괄적인 활용 가능성을 논의한 연구가 주를 이뤘다.

Lund와 Wang(2023)은 ChatGPT의 개념과 역사를 고찰하며, GPT와 AI를 도서관과 문헌정보학계에 적용할 수 있는 방법, ChatGPT를 이용함에 따르는 윤리적인 이슈나 저작권 문제, 도서관에서 자연어처리를 위해 ChatGPT를 사용할 때의 장점이나 의미 등을 실제로 ChatGPT와 인터뷰한 결과를 수록해 그 실례를 제공하였다. 이 결과를 통하여 이들은 ChatGPT가 매우 정교한 챗봇이며, 텍스트 기반의 질의에서 매우 강력한 능력을 보유하고 있음을 확인하였다. 그러면서 이 기술이 학계나 도서관계에서 양날의 검으로 작용할 가능성이 높으므로, 보다 책임감 있고 윤리적으로 사용해야 할 방법을 모색해야 할 것임을 주장하였다.

Cox와 Tzoc(2023)는 연구, 정보서비스, 교육, 인쇄 책자, 정보리터러시, 창작, 표절, 저작권, 생산성, 형평성과 포용성 등 도서관에서 수행하는 역할이나 기능에 적용하여 ChatGPT가

보유한 장점을 파악하였으며, 특히 ChatGPT의 대화적 요소는 연구에서의 초록을 생성하고 목록 추출, 인용 작성, 연구 과정의 단순화, 데이터 분석 등에서 두각을 나타낼 것이며, 최상의 생산성을 보일 것으로 예상하였다. 그러나 ChatGPT가 과금 형식을 적용하면서, 이용자들에게 불평등을 야기시킬 것이며 더불어 저작권이나 표절 문제에서도 신뢰성 결여 및 공정성에 대한 문제가 생성될 것을 우려하였다. 한편, 도서관으로서는 이용자들의 즉답형 질문에 대한 답변이나 24/7 서비스 등은 ChatGPT로 대신하거나, 정보리터러시나 교육 부문에서도 질문과 답변의 형식에 ChatGPT를 활용함으로써 사서의 업무를 지원할 수 있을 것으로 예상하였다. 이들은 이처럼 AI 도구가 인간의 삶의 개선을 도모할 잠재력이 있음을 인정하면서도 동시에 사서직군에 미칠 영향을 예측하기는 매우 어렵다는 것을 피력하고, 도서관에서는 AI 도구를 평가하고 지원하는 서비스를 개발해 AI 기술을 통제할 수 있어야 한다고 하였다.

Kirtania와 Patra(2023)는 앞의 두 선행연구에서 ChatGPT가 지닌 신뢰성 결여와 표절성을 식별하기 위하여 문헌정보학 분야의 주요 10대 주제(문헌정보학 일반, 공공도서관, 계량정보학, 디지털도서관, 도서관경영, 정보검색, 학술 커뮤니케이션, 인용 분석, 표절, 연구데이터 관리)를 선정하고 ChatGPT에게 이와 관련한 콘텐츠를 제작하게 한 후, 표절 분석 프로그램 'Turnitin'을 통해 답변과의 유사성을 확인하였다. 그 결과 전체 유사성은 13% 이하로 나타나 표절성이 낮은 것으로 확인되었으나, 결과물의 효용성을 평가하기 위해서는 더 많은 실험과 연구가 필요하다고 제안하였다.

본 연구는 문헌정보학 분야에서 메타데이터 관련하여 ChatGPT를 활용한 시도적 연구로서 연구자들의 연구 확장을 위한 기반으로의 역할을 기대할 수 있을 것이다.

3. 이론적 배경

3.1 더블린 코어

1995년에 OCLC(Online Computer Library Center)와 NCSA(National Centre for Supercomputer Application)는 미국 오하이오주 더블린에서 열린 워크숍에서 메타데이터 관련 합의를 이뤄냈고, 이를 기념하기 위해 개최지의 이름을 따서 더블린 코어(이하 DC)라고 부르게 되었다. 이치주 외(2000)는 더블린 코어의 이점으로 데이터 구조의 단순성, 이용의 용이성, 색인 작성의 단순화, 상호운용성 확보, 자원의 접근과 확보의 용이, 자원 관리의 질적 개선을 손꼽았다. DC는 15개의 기본요소와 그 외의 확장요소(qualified elements)로 나뉘는데, 그중에서 15개 기본요소에 대한 명칭 및 설명은 <표 1>과 같다.

이 표에 따르면 Subject와 Coverage처럼 일반적인 의미와 달리 더블린 코어에서는 의도된 특수한 목적을 처리하기 위해 사용되는 단어가 있다. 마찬가지로, 국내 도입 과정에서 더블린 코어의 각 요소 이름과 그 설명이 출처마다 조금씩 다르다는 사실에 주목할 필요가 있다. 이는 언어모델을 사용할 때 맥락상 잘못된 오류(예, 주제와 요약을 같은 개념으로 이해하는 등)로 귀결될 수 있기 때문이다.

〈표 1〉 더블린 코어 15대 요소

영문 요소 이름	국립중앙도서관 기준 ¹⁾		이치주 외(2000) 기준		의미상 호환 가능한 이름
	요소 이름	설명	요소 이름	설명	
Title	표제	자원에 부여된 제목	제목	제작자(creator)나 발행자(publisher)가 자원에 부여한 제목	
Creator	창작자	자원의 내용에 주된 책임을 가진 개체	저자	제작자(creator)나 발행자(publisher)가 자원에 부여한 제목	글쓴이
Subject	주제	자원의 내용적 주제(topic)	주제	자원의 주제나 그 내용을 기술하는 키워드 혹은 구절	주제명
Description	설명	자원의 내용에 대한 설명	기술	문서의 이미지의 요약 정보를 포함한 자원의 내용에 관한 정보 (예: 문헌의 초록, 시각자료의 내용기술 등)	요약
Publisher	발행자	자원을 현재의 형태로 이용 가능하게 만든 실체	발행자	자원을 현재의 형태로 이용 가능하게 만든 실체 (예: 출판사, 대학, 기업체)	출판사, 퍼낸 곳
Contributor	기여자	제작자 요소에 명시된 개체 이외에 자원의 내용에 기여한 책임이 있는 기타 개체	기타 기여자	저자 이외에 자원의 지적인 측면에 기여한 인물이나 기관(예: 편자, 번역자, 삽화가 등)	
Date	날짜	자원의 존재 기간 동안 어떠한 사건이 발생한 날짜, 자원의 제작일 또는 자원이 현재의 형태로 이용 가능하게 된 시점과 관련됨	발행일	자원이 현재 형태로 이용 가능하게 된 날짜	발행연도, 출판시기
Type	유형	자원의 내용이 가지는 성격 또는 장르로, 내용의 일반적인 범주, 기능, 장르 등을 표현함	유형	자원의 범주나 장르 (예: 홈페이지, 소설, 시, 토의문서, 기술보고서 등)	장르
Format	형식	자원의 물리적 표현형식 및 디지털 표현형식	표현방식	자원의 데이터 표현 형식 (예: text/html, ASCII, PostScript 등)	형태
Identifier	식별자	자원을 식별하기 위한 식별기호로 공인된 식별체계(URI, URL, DOIS, ISBN 등)를 따르는 문자/숫자열을 사용할 것을 권장함	식별기호	다른 자원과의 관계, 공식적인 관계를 가지면서 독립적으로는 존재하는 자원들과의 관계 표현	
Source	출처	현재 자원의 출처가 되는 원 정보 자료로의 참조	출처	해당 자원의 출처가 된 정보자원 (예: source(scheme=ISBN)=0-201-63337-X)	
Language	언어	자원의 지적인 내용을 기술하고 있는 언어	언어	자원의 내용을 기술한 언어 (예: 영어, 한국어, 독일어 등)	
Relation	관련자원	관련 자원에 대한 참조	관계	다른 자원과의 관계, 공식적인 관계를 가지면서 독립적으로는 존재하는 자원들과의 관계 표현	
Coverage	수록범위	자원의 내용이 다루는 범위로 공간적 위치나 시간적 범위(시대)가 해당됨	범위	자원의 지리적, 시간적 특성을 나타내는 요소	시공간 배경
Rights	이용조건	자원이 가지고 있는 권리나 자원에 대한 권리에 관한 정보	권리	저작권의 사용 권한에 관한 내용, 판권 사항이나 판권관리사항, 혹은 이러한 정보를 다양한 방식으로 제공하는 서버와 연결	저작권

1) 국립중앙도서관이 제공하는 더블린 코어 메타데이터의 15개 요소에 대한 설명.
<https://librarian.nl.go.kr/LI/contents/L10202000000.do>

Guinchard(2002)는 현실에 존재하는 객체를 담기에는 더블린 코어 요소의 수가 너무 적지만 사용 지침은 추상적이라는 특징에 착안해, 더블린 코어를 사용하고 있는 29개 도서관의 현황을 조사하였다. 이들 도서관은 더블린 코어를 사용하는 이유를 이 표준이 갖는 국제적 인지도와 유연성 때문이라고 답했다. 조사 대상의 9%만이 더블린코어 15개 요소 원칙을 그대로 사용하고 있고, 73%는 기관 고유의 한정자(qualifier)를 추가해서 사용하고 있었다. 더블린코어 15개 요소 중에서 가장 많이 쓰이는 것은 Creator, Title, Rights, Publisher, Type, Identifier, Description, Format으로 확인되었다. 이경호(2011)는 DC의 Description, Type, Format, Source, Language, Relation, Coverage, Rights 등의 요소는 도서 자원보다는 인터넷 자원의 특징을 나타내는 더 적합할 수 있다고 평가하였다.

3.2 메타데이터 품질측정

Park(2009)는 메타데이터 품질 기준의 3대 요소로써 동일한 메타데이터 표준을 사용하더라도 분야마다 필수요소(mandatory elements)와 선택요소(optional elements)는 다를 수 있다는 사실을 반영한 완전성(completeness), 자원의 올바른 내용을 담고 있는지에 대한 정확성(accuracy), 자원을 표현하는데 동일한 표현양식과 용어를 사용하고 있는지에 관한 일관성(consistency)을 언급하였다. Ochoa와 Duval(2009)은 메타데이터 자동 생성과 관련해 품질을 평가하는 측정방안으로 완전성, 정확성, 일치성(conformance), 일관성(consistency

& coherence), 접근성(accessibility), 적시성(timeliness), 기원성(provenance)을 도출하기 위한 계산식을 정리하였다. Johnson et al.(2023)은 ChatGPT가 생성한 의학적 의견을 평가하는 방법으로 전문가에게 직접 정확성과 완전성의 수준에 대한 점수를 부여하도록 하였다.

품질측정의 맥락에서 자원 유형별로 진행된 연구를 살펴보면 다음과 같다.

웹 자원의 경우 더블린 코어는 소수의 기술 요소를 가지면서 몇몇 요소가 비교적 포괄적이거나 범용적인 의미를 가지므로 메타데이터 작성자의 전문가적 지적 판단이 매우 중요하다. 이를 실제로 사후 평가한 결과 더블린 코어가 내포하고 있는 개방성에 기인한 다양한 부정확성 문제가 확인되었다(Sokvitne, 2000; Zavalina & Burke, 2021; Zhang & Dimitroff, 2004; 2005a; 2005b).

도서의 경우 Chapman과 Massey(2002)는 영국 Bath 대학교 도서관의 MARC 데이터 중 일부를 표본 추출하여 검증하는 과정에서 34.4%가 최소한 하나의 오류를 포함하고 있음을 확인하였다. 그 중의 중요 오류만 하더라도 7.6~11.4%에 이르렀고, 5.9%는 다수의 필드에서 오류가 발생하였으므로 오류 발생 패턴에 군집적 특성이 있을 것이라고 예상하였다. James와 Weiss(2012)는 Google Books의 도서 메타데이터를 평가함으로써 내재한 오류 규모가 어느 정도인지 파악하고자 하였다. 400권의 표본 데이터를 검증한 결과, 36.75%의 메타데이터 오류가 발견되었고 그 빈도와 비율순으로는 Publisher, Author, Publication Date, Title이었다.

학술지의 경우 이용구, 김병규(2010)는 새로운 메타데이터 품질 계량화 방법을 제안하며

정보연구원(KISTI)의 과학기술학회마을에서 제공하는 학술지 기사 메타데이터 총 974건을 평가하였다. 학술지는 웹 자원이나 도서와 달리 메타데이터 오류율이 매우 낮았다. 완전성 측면에서의 오류는 이메일 주소 누락이 대부분이었으며, 정확성 측면에서의 오류는 주로 외국어 키워드와 초록에서의 오타자였다.

3.3 메타데이터 자동 추출 및 생성 연구

추출과 생성은 상호보완적인 개념으로, 가용한 정보원으로부터 적합한 정보를 취득한 것을 추출이라 하며 추출된 정보를 정제하거나 합성함으로써 새로운 차원의 정보를 만드는 것을 생성이라 한다(Huang et al., 2022, Valls-Vargas, 2013). 이 관점에서 용성중 외(2021)는 영상 자원에서 소리를 추출한 뒤 해당 자원의 장르와 언어에 대한 메타데이터 요소를 생성하는 기법을 제안하였다.

Underwood(2020)는 미국 백악관 보도실 자료부터 더블린 코어를 자동 추출하는 연구를 진행하였는데 Subject, Description, Contributor는 요소는 빠진 상태였다. 이처럼 추출할 수 있는 메타데이터 요소가 전혀 없는 데이터인 경우, 당연히 생성 과정을 거치더라도 빈 채 남아 있을 수 있다.

양기철, 박정란(2018)이 제안한 새로운 자동화 메타데이터 추출 기법은 입력된 자연어를 개념 그래프(conceptual graph)로 변환함으로써 문장에서 분리된 단어의 의미가 모호해진 결과 메타데이터 요소가 품질이 저하되는 것을 예방하는 데 목적을 두었다. 이 과정의 예시로서 DC의 Creator, Title, Contributor 추출 규칙을 설

명하였다. Ojokoh, Adewale, Falaki(2009)는 인터넷에 공개된 40개의 PDF의 학술자료를 대상으로 제안한 규칙 기반 메타데이터 추출 모델을 적용하고 그 성능을 정확률(precision), 재현율(recall), F1, 정확성으로 측정함으로써 자동 메타데이터 생성의 가능성을 제시하였다. Han et al.(2003)은 컴퓨터공학 분야 연구논문의 첫 번째 페이지에 있는 정형화된 헤더 데이터를 수집하여 SVM을 사용하여 메타데이터를 추출한 결과 전체 정확성이 92.9%에 달하는 높은 성능을 보였다. 정종진, 김경원, 김구환(2020)은 입력된 텍스트 데이터로부터 빈도분석과 토픽모델링을 사용해 키워드와 주제어를 추출한 뒤 메타데이터를 생성하는 모델을 제안하였다. Irvin(2003)은 당시 널리 사용되던 UC-Dot과 Klarity와 같은 웹페이지용 더블린 코어 자동 생성 도구를 통해 구축된 메타데이터와 전문 인력이 생성한 메타데이터의 품질을 비교 평가하였으며, 전문 인력이라고 해서 자동 생성 도구보다 더 현저한 고품질의 메타데이터를 만드는 것은 아니라고 결론지었다. 따라서 오늘날과 같이 대용량의 데이터로부터 자동으로 메타데이터를 생성할 수 있는 기술적 요구와 성숙도가 모두 갖추진 지금이야말로, 자동 메타데이터 도구의 가치가 그 어느 때보다 높다고 할 수 있겠다.

3.4 GPT와 ChatGPT

정확도가 높은 언어모델을 만들기 위해서는 되도록 많은 레이블 데이터(labeled data)를 학습하는 것이 중요하다는 것이 오랜 통론이었으나, 대부분의 연구자 관점에서 이는 여전히 해

결하기 어려운 일이다. 이를 해결하기 위해, 먼저 레이블이 없는 대용량의 데이터 세트를 이용하여 비지도 학습한 뒤 소량의 레이블 데이터로 파인튜닝(fine-tuning) 하는 전이학습 지원 모델들이 등장하였다. 이런 배경으로 등장한 언어모델이 최초의 GPT 모델인 GPT1과 BERT이고, 이 과정에서 핵심적인 역할을 하는 것이 트랜스포머(Transformer)이다(Devlin et al., 2018; Radford et al., 2018). 두 모델은 비슷한 시기에 유사한 접근법으로 구현된 것이지만, 단방향으로 학습함으로써 문장 생성 능력에 강점을 지닌 GPT1보다 양방향으로 학습함으로써 문장 의미추출 능력에 유리한 BERT의 성능이 종합적으로 높았을 뿐더러 파인튜닝에 필요한 자원이 Hugging Face를 통해 적극적으로 공개되었으므로 연구자와 개발자는 BERT를 선호하는 경향이 더 높았다.

이후 등장한 GPT2의 가장 큰 특징은 파인튜닝 과정을 배제함으로써, 준지도학습에서 비지도 학습 모델로 특성이 바뀌었다는 데 있다(Radford et al., 2019). 즉, 비지도 학습 모델로부터 다운스트림 작업(down-stream task)에 적합한 결과를 제로샷(zero-shot) 방식으로 제공할 수 있다. 이 모델은 4백만 개의 하위 링크로 구성된 텍스트 데이터셋인 WebText를 학습하였는데, 이는 8백만 건 이상의 문서에 해당하는 분량이며 텍스트파일로 환산하면 그 크기가 약 40GB에 달한다. 학습 모델 구축 시, BPE(Byte Pair Encoding) 방식을 채택하여 자주 사용되지 않는 어휘도 처리할 수 있도록

하였으며 입력 항목으로 입력 텍스트 외에도 태스크 라벨을 반영함으로써 다양한 형태의 태스크에 대응할 수 있는 범용 모델을 추구하였다. 비영어권 데이터인 중국어로 글쓰기 성능을 측정한 사례에서도, 특성 차이는 있으나 BERT와 GPT2 모두 인간이 읽는 데 문제없는 수준의 성능을 나타내었다(Qu et al., 2020).

GPT3은 GPT2와 동일한 구조와 모델을 사용하지만, 트랜스포머 계층 사이에 몇 가지 변경 사항(alternating dense와 locally banded sparse attention)을 추가하고 최대 학습 파라미터 수를 기존의 15억 개에서 1,750억 개로 늘리고, 웹에서 크롤링 된 문서를 학습하였다. 결과적으로 영어 이외의 언어에 대해서도 학습한 데이터를 가지고 있으며, 한국어의 경우 단어 기준 29위로 데이터 세트의 0.01697%를 차지²⁾한다. Armengol-Estapé, Bonet, Melero(2021)는 GPT3의 소수 언어 생성 능력을 파악하기 위해 학습데이터 내에서 0.01798%만을 차지하는 카탈로니아어³⁾를 대상으로 Q&A 태스크에 대한 평가를 시행하였다. 실제 카탈로니아어를 쓰는 사람이 쓴 답변을 통제 변인으로 하고 GPT3가 쓴 답변과 비교한 결과, 물론 인간의 글쓰기 점수(4.49)가 더 높았지만, GPT3의 성능(3.83)도 상당한 수준이었으며 일부 평가항목에서는 인간보다 GPT3의 글쓰기가 더 나은 점수를 얻었다. 이처럼 다재다능한 GPT3이지만, 전문 분야에서 전용 데이터로 파인튜닝된 BERT 기반 언어모델과 비교하면 낮은 성능을 보일 수도 있다(Moradi et al., 2021).

2) https://github.com/openai/gpt-3/tree/master/dataset_statistics

3) 단어 기준, 카탈로니아어는 28위이며 한국어는 29이다.

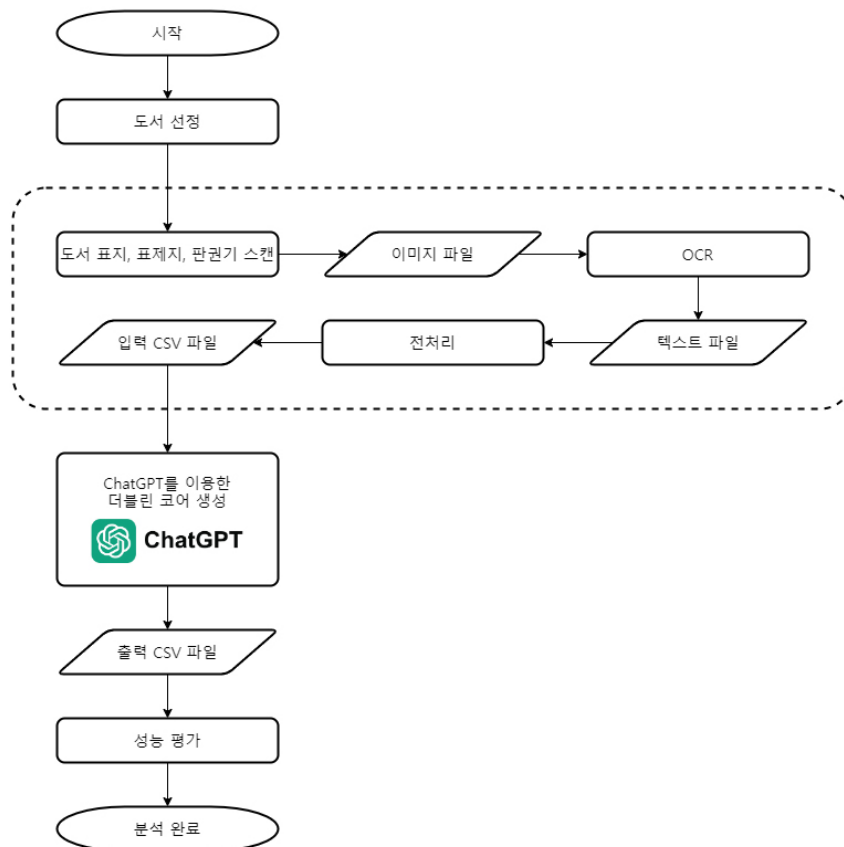
OpenAI사는 이후 GPT3 시리즈의 개선 버전인 GPT3.5를 선보였으며, 이 중에서 text-davinci-003 모델이 가장 높은 성능을 보이나 경제성과 사용성 측면에서 gpt-3.5-turbo 모델이 권장⁴⁾된다.

ChatGPT는 OpenAI사가 GPT3 시리즈와 GPT4 시리즈 언어모델을 웹 페이지 상에서 챗봇 형식으로 구현한 것이며, 웹 인터페이스 이외에도 개발자를 위한 API를 제공하고 있다.

4. 연구방법

4.1 연구 설계

본 연구는 <그림 1>과 같이 4개 절차로 구성된다. ChatGPT에게 '사서'와 '메타데이터 전문가'라는 역할을 부여한 뒤, 사서가 실제로 입수한 도서에 대하여 메타데이터를 부여할 때 수행하는 전문가적 지적 작업을 ChatGPT가 수행하도록 그 과정을 유도하였다.



<그림 1> 연구 절차

4) <https://platform.openai.com/docs/models/model-endpoint-compatibility>

첫 번째, 본 연구에 사용될 대상 도서의 범위를 선정하고 확보하였다.

두 번째, 대상 도서에서 메타데이터 작성에 필요한 실제 인쇄된 내용을 스캔하고 이를 컴퓨터가 처리할 수 있도록 다시 문자 데이터로 변환하였다.

세 번째, 대상 도서로부터 최종 확보된 텍스트 데이터를 ChatGPT에 입력하고 더블린 코어 작성을 요구한 뒤 그 결과를 확보하였다.

네 번째, ChatGPT가 작성한 더블린 코어가 어느 정도 충실하고 정확한 내용을 담고 있는지 파악하기 위해 이를 평가하였다.

4.2 데이터 수집 및 전처리

본 연구는 K대학교 도서관에서 이용 빈도가 높은 도서군(圖書群)을 대상으로 층화표집(stratified random sampling) 방식에 따라 <표 2>와 같이 '인문사회과학', '과학기술', '문학' 분야별 30권씩 총 90권의 도서를 선정하여 확보하였다.

선정된 도서를 실제 살펴본 결과, 메타데이터에 사용될 수 있는 정보가 인쇄된 도서의 구성 요소로 표지(front cover), 표제지(title page),

판권기(colophon)가 식별되었으며 이는 국립중앙도서관의 온라인자료 메타데이터 DB 구축 지침⁵⁾과도 일치하였다.

전처리 과정으로는 우선 사서가 메타데이터를 부여하기 위해 도서의 정보를 획득하는 과정에 사용하는 도서의 표지, 표제지, 판권기를 스캔하여 이미지로 확보하였다. 그다음 사서가 시각적으로 인식한 정보를 자신의 전문 지식을 바탕으로 메타데이터 작성하는 과정을 ChatGPT가 수행할 수 있도록, 이미지 파일을 광학문자인식(OCR) 처리하고, 이를 전처리한 후 CSV 파일로 정리하였다. 광학문자인식을 위해 네이버 AI 서비스(NAVER Cloud for AI)를 이용하였고, 이후 적용된 전처리 규칙은 <표 3>과 같다.

4.3 ChatGPT를 사용한 더블린 코어 생성

프롬프트 엔지니어링(prompt engineering)은 GPT3와 같은 대규모 언어모델로부터 적합한 결과를 도출하기 위한 접근법을 의미하는 용어로, 관련된 연구가 활발히 진행되는 추세이다(Liu & Chilton, 2022; White et al., 2022; Zhou et al., 2022). OpenAI사는 이용자의 입

<표 2> 도서 선정 현황

분야	DDC 주류	권 수 할당	
		인문사회과학	과학기술
인문사회과학	철학(100)	15권	30권
	사회과학(300)	15권	
과학기술	자연과학(500)	15권	30권
	기술과학(600)	15권	
문학	문학(800)	30권	30권

5) <https://oak.go.kr/nl-ir/handle/2020.oak/664>

〈표 3〉 전처리 규칙

구분	전처리 규칙
표지 표제지	1. 입력순서: 인쇄 내용 기준 상->하, 좌->우. 단 순서가 모호할 경우, 제목-저자-출판사 순서 입력 2. 내용: 부제 및 책 소개 문구 포함
관련기	1. 저자와 옮긴이의 약력: 제외 2. 출판사 소개 등 홍보 문구: 제외 3. 출판시도서목록(CIP) 제외
공통	1. 주요 특수문자(· 및): “·”는 문맥에 따라 삭제하거나 “,” 또는 “ 및 ”으로 치환 는 빈칸으로 치환 2. 심미용 특수문자 삭제 3. 판독 불가능한 문구 삭제 4. 한국어 외 표기 제외

력을 프롬프트(prompt)라고 칭하며, 이용자에게 반환된 출력 결과를 컴플리션(completion)이라고 정의한다. 프롬프트는 크게 시스템 프롬프트, 어시스턴트 프롬프트, 유저 프롬프트로 나눌 수 있다. 시스템 프롬프트는 어시스턴트의 행동을 설정하는 역할을 하며, 어시스턴트 프롬프트는 그 내용을 기억하여 이후에도 사용할 수 있도록 하는 부분으로 ChatGPT가 스스로 답변을 이어갈 때 사용하거나 개발자가 ChatGPT에 기억해야 할 내용을 전달할 때 사용한다. 유저 프롬프트는, 복수의 프롬프트로 구성된 하나의 애플리케이션의 말미에서 ChatGPT에게 명령이나 지시를 전달하는 역할을 한다.

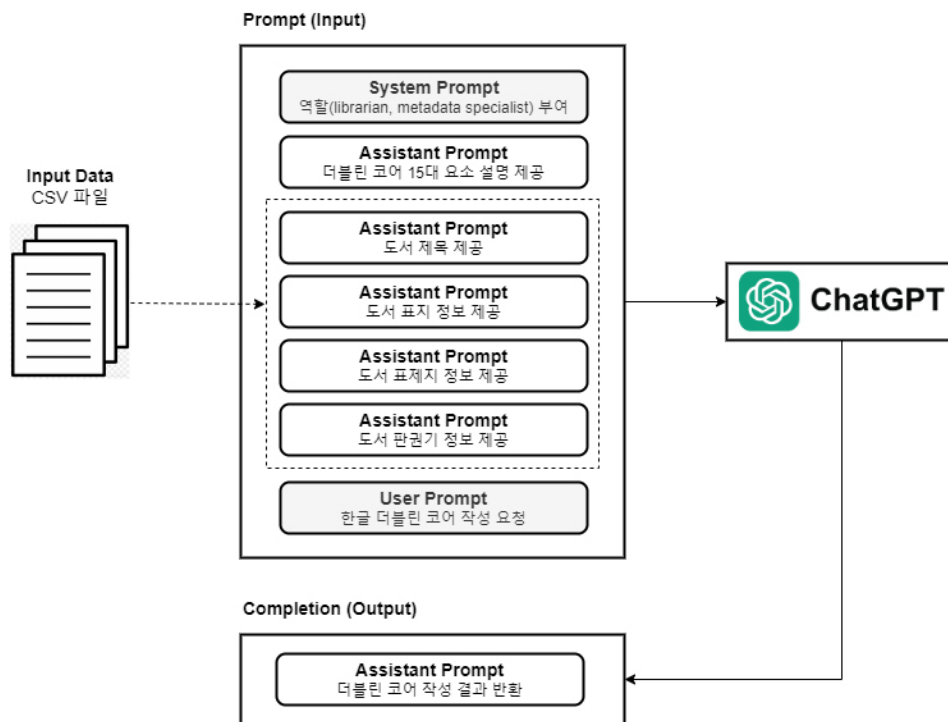
본 연구는 ChatGPT에게 최대한의 예시를 제공하여 더블린 코어를 만들기보다는 최소한의 정보만으로도 더블린 코어를 작성하는 것에 초점을 두었다. 특히, “지시[질문]를 명확하게 작성하고, 참조정보 또는 예제를 제공하라”⁶⁾는 ChatGPT의 권고를 따라 결과의 품질을 보장되도록 하되, 인간의 개입을 최소화하도록

설계하였다. 〈그림 2〉는 이런 과정을〈그림 1〉의 ‘ChatGPT를 이용한 더블린 코어 생성 과정’을 구체적으로 나타낸 것으로, 먼저 시스템 프롬프트를 통해 ‘사서’와 ‘메타데이터 전문가’의 역할을 부여하고, 어시스턴트 프롬프트를 통해 CSV 파일로 수집해둔 도서 정보를 전달하고, 마지막으로 유저 프롬프트에서 더블린 코어 작성을 명령하였다.

프롬프트를 ChatGPT에 입력하고 출력된 결과인 컴플리션을 저장하는 과정을 같은 세션에서 사람이 직접 입력할 경우, 이전 입출력 결과가 이후 입출력 결과에 영향을 미칠 가능성이 있으며 더불어 수행시간이 상당히 소요된다. 따라서 본 연구에서는 파이썬에 기반하여 OpenAI ChatGPT API를 사용한 ChatGPT 입출력 프로그램을 제작하였다.

본 연구의 실험에 사용된 언어모델은 gpt-3.5-turbo이며, 하이퍼파라미터인 temperature는 기본값인 0.7, maximum length는 최댓값인 2048로 설정하였다.

6) <https://platform.openai.com/docs/guides/gpt-best-practices>



〈그림 2〉 ChatGPT를 이용한 더블린 코어 생성 과정

여기서 temperature는 ChatGPT의 창의성을 통제하는 값으로 1에 가까울수록 단어 선택의 다양성이 높아진다. 즉, ChatGPT에게 매번 다른 창의적인 답변을 원하고 싶다면 1에 가깝도록 설정하는 것이, 반대로 매번 같은 답변을 받을 확률을 높이고 싶다면 0에 가까운 낮은 값으로 설정한다. 본 연구에서는 기본값인 0.7 설정을 사용하였다. maximum length는 사용자의 애플리케이션이 여러 프롬프트를 통해 입력하고 출력할 수 있는 전체 토큰 개수가 의미한다. 토큰의 길이는 입출력하는 언어에 따라 다른데, 영어는 1개의 토큰이 최대 4개의 문자에 해당하며 한글은 3개의 토큰이 1개의 문자에 상응한다.

4.4 성능평가

메타데이터의 품질을 평가하는 방법은 필요한 메타데이터 요소가 모두 다 작성되었는지에 대한 완전성과 각 요소가 올바르게 기술되었는지에 대한 정확성을 측정하는 것이 일반적이고, 연구에 따라 정성적인 방법과 정량적인 방법 둘 중 적절한 것을 선택할 수 있다(이용구, 김병규, 2011; Margaritopoulos et al., 2012; Ochoa & Duval, 2009).

이에 따라 본 연구는 ChatGPT의 더블린 코어 생성에 관한 완전성과 정확성을 측정하였다. 이 두 평가척도를 식으로 나타내면 다음과 같다. 이 식에서 N 은 메타데이터 스키마에서 받

드시 기술해야 하는 요소의 개수이다.

$$Completeness = \frac{\sum_{i=1}^N P(i)}{N} \quad \text{수식 (1)}$$

$$Accuracy = \frac{\sum_{i=1}^N H(i)}{N} \quad \text{수식 (2)}$$

수식 (1)은 완전성을 구하는 방법이다. 식에서 $P(i)$ 함수는 각 요소가 등장하였으면 1을 그렇지 않으면 0을 부여한다. 따라서 $\sum P(i)$ 는 등장한 요소의 개수이다.

수식 (2)는 정확성을 구하는 방법이며, $H(i)$ 는 실제로 각 요소가 적합할 경우 1을 그렇지 않으면 0을 부여한다. 따라서 $\sum H(i)$ 는 적합한 요소의 개수이다. 적합 여부는 정사서 1급 자격증 보유자 1인과 정사서 2급 자격증 보유자 1인의 의견이 일치 여부로 판정하였다.

한편, 원본에 나타난 형태 그대로는 메타데이터로 사용될 수 없는 데이터를 가공하여 메타데이터 요소를 작성하는 행위를 생성이라 할 수 있다(용성중 외, 2021). 따라서 본 연구에서 더블린 코어 생성이란 도서의 표지, 표제지, 판

권기에 실린 정보를 중심으로 어떤 도서의 '데이터의 데이터'를 구조화된 형태로 구축하는 행위이자 사서가 지닌 배경지식과 지적 사고가 깊게 관여된 지식 활동으로 정의한다. ChatGPT는 사서의 이러한 전문 지식 활동과 유사한 수준으로 더블린 코어를 생성할 수 있어야 한다.

이명규(2010)는 판권기에 기재된 기술서지 요소로 저작자(저자 및 역자), 도서명, 인쇄, 발행, 등록 및 발행인, 정가, 국제표준도서번호(ISBN), 출판시도서목록(CIP), 저작권 표시가 있다고 하였다. 이 중 더블린코어 요소에 포함할 수 있는 기술서지 요소는 <표 4>와 같이, 정가와 출판시도서목록(CIP)를 제외한 기술서지 요소 7개로 나타났다. 하지만 인쇄는 인쇄일을 발행은 발행일을 의미하고 있으므로, 더블린코어 요소에서 날짜를 작성할 수 있는 Date로 포함하여 결과적으로 판권기에서 확인할 수 있는 더블린코어 요소는 총 6개로 Title, Creator, Date, Publisher, Identifier, Right이다. 이에 따라 판권기뿐만 아니라 도서의 표지, 표제지의 내용을 포함하여 교차 분석한 결과, 다른 정보원의 도움을 받지 않고 각 도서가 포함하고 있는 정보만으로 사서가 생성할 수 있는 더블

<표 4> 판권기 기술서지 요소와 DC 요소의 연결

판권기 기술서지 요소	더블린코어 요소
저작자(저자 및 역자)	Creator
도서명	Title
인쇄	Date
발행	Date
등록 및 발행인	Publisher
정가	-
국제표준도서번호(ISBN)	Identifier
출판시도서목록(CIP)	-
저작권	Rights

린 코어 요소는 제한적이라고 판단하였다.

한편, <표 5>는 본 연구의 입력데이터(표지, 표제지, 판권기)로부터 사서 또는 ChatGPT가 유추하여 생성할 수 있다고 예측되는 더블린 코어 요소의 후보군을 사례별로 구분한 것이다. Subject, Contributor, Description의 경우, 판권기에 기록된 요소는 아니지만, ChatGPT가 사서처럼 유추할 수 있다면 이 요소를 작성할 수 있을 것으로 기대하였다.

본 연구에서 사용된 도서 대부분이 사례 1 (Source, Relation, Coverage를 유추할 수 있는 정보가 판권기 내 기재되어 있지 않으며, 이를 유추하려면 외부 정보원이 필요한 경우)에 해당하였으나, 사례 2(사례 1 + 판권기 내 Contributor 정보가 없는 경우)와 사례 3(사례 1 + 판권기 내 Rights 정보가 없는 경우)과 같은 예외 사례도 확인되었다. 즉, 더블린 코어 생성에 필요한 서지 데이터 분포 현황은 도서마다 다르므로

ChatGPT가 생성한 메타데이터의 품질을 측정하기에 앞서 각 도서를 <표 5>의 사례별로 분류하였다. 그 결과 수식 (1)과 수식 (2)의 평가 척도에서 N 의 값은 도서에 따라 $N=11$ 또는 $N=12$ 가 되며 이를 통해 완전성과 정확성을 평가하였다.

5. 연구결과

5.1 전체 성능

ChatGPT를 사용하여 생성한 메타데이터 결과의 전체 품질 성능은 <표 6>과 같다.

완전성과 정확성 점수 평균으로 3단계(우수: 0.8 이상, 보통: 0.6 이상 0.8 미만, 미흡: 0.6 미만) 범주로 고려했을 때 완전성은 평균 0.87이며 정확성은 평균 0.71로 나타나, 완전성

<표 5> 더블린 코어 생성 가능 패턴

DC 요소	사례 1 ($N=12$)	사례 2 ($N=11$)	사례 3 ($N=11$)
Title	○	○	○
Creator	○	○	○
Subject	○	○	○
Description	○	○	○
Publisher	○	○	○
Contributor	○	-	○
Date	○	○	○
Type	○	○	○
Format	○	○	○
Identifier	○	○	○
Source	-	-	-
Language	○	○	○
Relation	-	-	-
Coverage	-	-	-
Rights	○	○	-

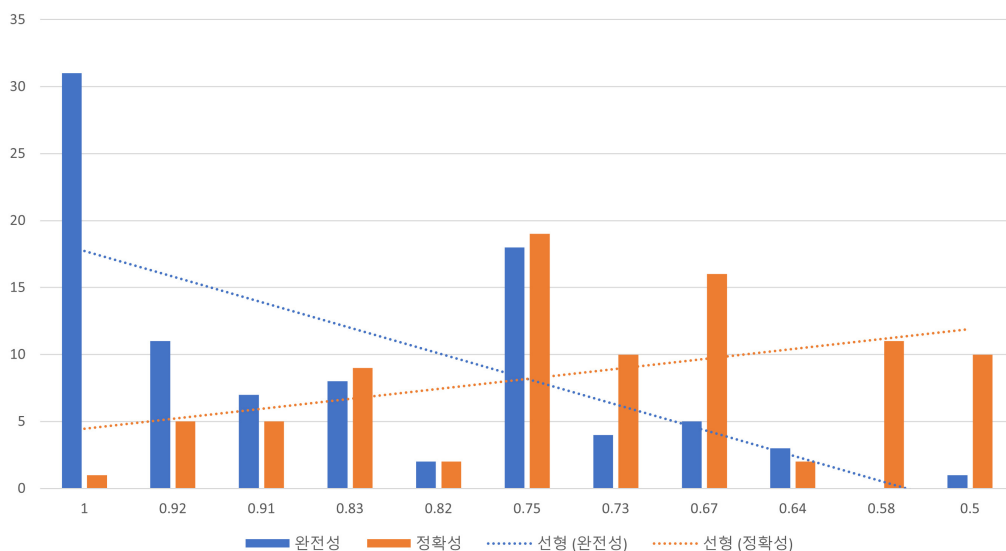
〈표 6〉 전체 데이터 기준 품질 평가

	평균	최대	최소	표준편차
완전성	0.87	1.00	0.50	0.13
정확성	0.71	1.00	0.50	0.12

과 정확성 모두 준수한 성능을 보인다고 판단하였다. 그러나 완전성의 경우 0.87 이상의 성능이지만 정확성은 0.71로 ChatGPT가 더블린코어 요소 생성에 있어 요소 자체의 추출은 비교적 알맞게 잘하고 있으나, 추출한 결과인 요소값이 실제로 올바른가에 대한 정확성은 완전성에 비해 상대적으로 낮은 것을 확인하였다.

전체 데이터에서의 완전성과 정확성 모두 최대 성능은 1이고, 최소 성능은 0.5로 성능 구간별 데이터 수를 살펴보면 완전성의 경우 전체 데이터 90건 중 66%에 해당하는 59건이 0.8의 성능의 이상을 나타냈으며, 0.8 미만에서 0.6 이

상의 성능은 33%에 30건이다. 정확성의 경우, 24%인 22건만이 0.8 이상의 성능을 보였으며 0.8 미만에서 0.6 이상의 성능은 52%에 47건이다. 성능 값별 데이터의 빈도수를 그래프로 그리면 〈그림 3〉과 같으며, 실험 데이터의 성능이 완전성일 경우 1, 정확성일 경우 0.75의 성능을 가장 많이 보이는 것으로 나타났다. 더불어 추세선으로도 확인할 수 있듯이 완전성이 정확성보다 1에 근접한 성능을 가진 사례가 매우 많기 때문에 성능이 높아질수록 매우 가파르며, 반대로 정확성은 성능이 낮아질수록 추세선이 높아지나, 그 각도가 상대적으로 완전성에 비하여 완만하였다.



〈그림 3〉 완전성과 정확성의 성능별 해당 데이터 수

한편 완전성에서 0.6 미만의 성능을 보인 사례는 〈그림 3〉에서 보이는 것처럼 단 1건으로, $N=12$ 에서 ChatGPT가 생성한 더블린 코어 요소의 개수는 6개에 불과했다.

5.2 더블린 코어 요소별 성능

ChatGPT가 생성한 더블린 코어 12개 요소 (Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Language, Rights)에 대한 성능은 〈표 7〉과 같다. 완전성과 정확성 점수 평균으로 3단계(우수: 0.8 이상, 보통: 0.6 이상 0.8 미만, 미흡: 0.6 미만) 범주로 고려하였다.

완전성에서 전체 평균인 0.87 성능보다 높은 성능을 보인 요소는 Title, Creator, Publisher, Date, Type, Format, Identifier, Language인 것으로 나타났으며, 특히 Title과 Creator는 성능이 1로 모든 데이터에서 생성된 요소로 확인되었다.

선행 연구에서 언급되었던 바와 같이 도서명, 저자명, 출판, 발행, 인쇄, ISBN, 저작권 등은 관련기 서지기술 내용에서 직접적으로 추출될 수 항목이고 ChatGPT도 이와 관련된 더블

린 코어 요소에 대해서는 높은 완전성을 보였다. Language의 경우 데이터에는 언어 기술 사항이 없었지만, 입력데이터가 한국어였기에 생성된 것으로 생각된다. 그리고 Type의 경우 ‘책, 도서, 서적’, Format의 경우 ‘인쇄물, 인쇄된 책, 출판물’ 등으로 ChatGPT가 실제 메타데이터를 생성하였는데, 이는 입력데이터가 이미 도서이고 인쇄물인 것을 ChatGPT가 인지하고 생성하였을 것으로 판단하였다.

한편 Subject, Description, Contributor는 전체 평균보다 낮은 완전성의 성능을 보였는데, 전체 90건의 데이터 중 Subject는 56건, Description은 55건, Contributor는 37건이었다. Subject와 Description은 사서의 지식 창출이 요구되는 요소로 ChatGPT가 완전성에서 각 0.62와 0.61의 성능을 나타낸 것은 전체 평균보다는 낮으나 제공된 도서 정보에 존재하지 않은 것을 이에 기반하여 새롭게 생성하고 있는 것을 확인할 수 있는 부분으로 ChatGPT의 장점이자 특성인 생성을 보여주는 결과라고 할 수 있겠다. 한편, Contributor는 저작물에 내용상으로 이바지한 자로 번역자 등을 포함한다. 그러나 ChatGPT는 도서에 기술된 저작자, 작가, 번역자, 삽화가 등 책임표시사항을 구체적으로 구분할 수 없어

〈표 7〉 요소별 완전성과 정확성 성능

요소	Title	Creator	Subject	Description	Publisher	Contributor
완전성	1.00	1.00	0.62	0.61	0.98	0.41
정확성	0.96	1.00	0.12	0.58	0.86	0.33
요소	Date	Type	Format	Identifier	Language	Rights
완전성	0.99	0.99	0.99	0.99	0.91	0.83
정확성	0.96	0.07	0.90	0.98	0.91	0.76

완전성 전체 평균: 0.87

정확성 전체 평균: 0.71

이들을 모두 Creator로 작성하는 등의 모습을 보여 Contributor의 완전성이 낮은 성능을 보이는 것으로 나타났다.

정확성에서 전체 평균 0.71보다 높은 성능을 보인 요소는 Title, Creator, Publisher, Date, Format, Identifier, Language, Rights였으며, 낮은 성능을 보인 요소는 Subject, Description, Contributor, Type이었다.

Title의 경우, 90건 중 4건이 부정확한 데이터를 생성하였는데, 예를 들어 ‘비만 코드’가 도서명이지만 ChatGPT는 표제지의 ‘체중은 인슐린이 결정한다.’라는 부제를 Title로 채택하여 생성하였다. 그리고 ‘이 책은 돈 버는 법에 관한 이야기’가 도서명인 데이터에 대해 ‘이 책은’을 제외한 ‘돈 버는 법에 관한 이야기’로 Title을 채택하여 생성하였다. 이러한 사례에 따라 ChatGPT는 표제지나 판권기에서 부제나 지시대명사 등의 구분을 다소 불확실하게 파악하는 것으로 보인다.

Subject와 Description 요소에서 완전성의 성능은 유사하였지만, 정확성에서 Subject가 현저히 낮은 성능을 보여주었다. Subject의 경우 ChatGPT가 국립중앙도서관의 주제명표목표를 사용하지 않았고, 자연어를 사용하더라도 어휘단위의 명사형보다 도서의 부제나 표제지에 기술된 광고 문구 등을 조합하여 생성했기 때문이었다. 반면 Description의 경우 Subject

와 유사한 방식으로 더블린코어의 요소로써 생성되었지만, 도서명이 ‘피로사회’처럼 매우 짧더라도 ChatGPT는 ‘현대인들의 일상에서 느끼는 스트레스와 피로에 대한 인식과 대처법을 다룬다’이라는 충실할 내용의 Description을 생성해냄으로써 ChatGPT의 특성을 확인할 수 있었다.

Contributor는 앞서 완전성에서 설명했던 바와 같이 ChatGPT가 저작의 책임표시사항에 대하여 구분하지 못한 결과가 동일하게 확인되었으며, Type은 더블린 코어의 장르 통제어를 전혀 고려하지 않고 요소값을 생성하여 완전성은 높았지만 정확성은 매우 낮은 값을 보였다.

요소별 성능을 요약하면 ChatGPT는 데이터 중 기술 요소가 명확히 나타난 요소에 대하여 우수한 생성 능력을 보여주었으며, 사서와 같은 지적 작업 활동이 요구되는 Subject나 Description 요소는 성능이 전체 평균보다는 다소 낮으나, ChatGPT 자체의 생성 역량을 확인할 수 있는 요소인 것으로 나타났다.

5.3 DDC 주류별 성능

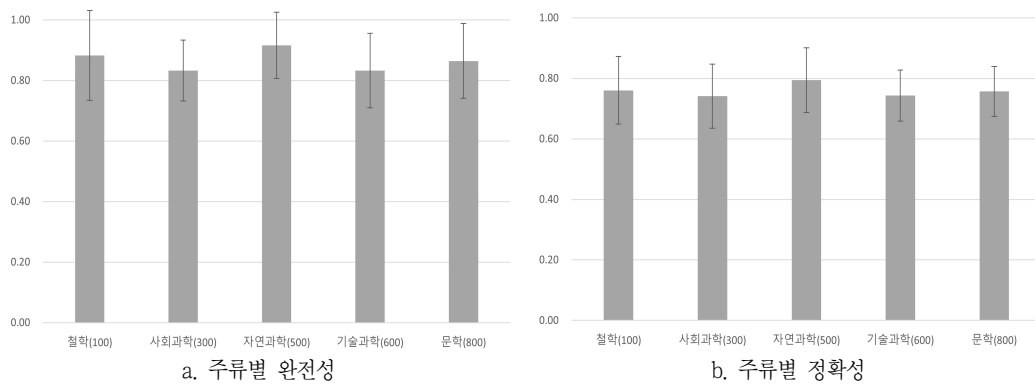
ChatGPT가 더블린 코어를 생성할 수 있도록 입력된 서지정보를 DDC 주류별로 구분하여 완전성 및 정확성 성능을 확인하면 <표 8>과 <그림 4>와 같으며, 완전성과 정확성 측면에서

<표 8> 주류별 완전성과 정확성 평균

요소	철학(100)	사회과학(300)	자연과학(500)	기술과학(600)	문학(800)
완전성	0.88	0.83	0.92	0.83	0.86
정확성	0.72	0.68	0.76	0.68	0.68

완전성 전체 평균: 0.87

정확성 전체 평균: 0.71



〈그림 4〉 주류별 데이터의 완전성 및 정확성

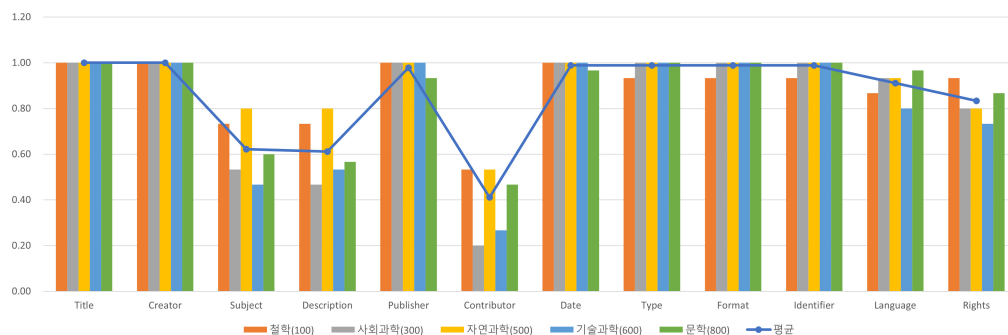
사회과학(300)과 기술과학(600)은 다른 주류에 비해 다소 낮은 성능을 보였다.

이에 따라 〈그림 5〉와 〈그림 6〉은 DDC 주류별로 요소의 성능을 시각화한 것으로, 완전성과 정확성 점수 평균으로 3단계(우수: 0.8 이상, 보통: 0.6 이상 0.8 미만, 미흡: 0.6 미만) 범주로 고려하였다.

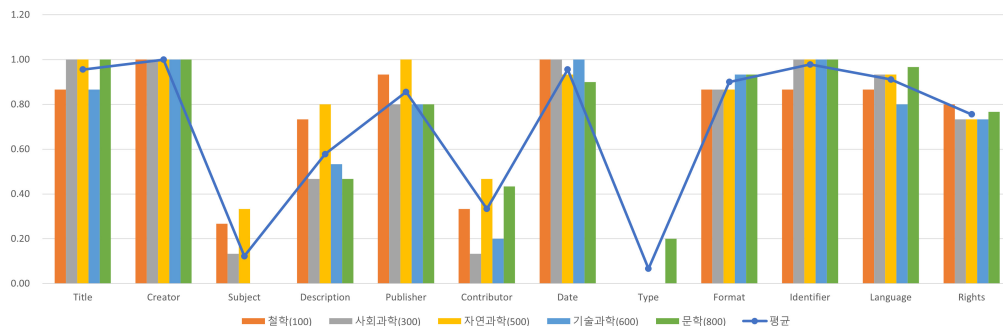
완전성 측면에서 ChatGPT는 〈그림 5〉와 같이 9개 요소(Title, Creator, Publisher, Date, Type, Format, Identifier, Language, Rights)에 대하여 주류 전반에서 우수한 성능으로 더블린 코어를 생성했다. 나머지 3개 요소의 경

우, Subject와 Description은 각각 0.62와 0.61로 보통이었고 Contributor 성능은 0.41로 가장 낮았다. 성능이 낮은 3개 요소 중에서 사회과학(300)과 기술과학(600) 분야는 다른 분야에 비해 더 낮은 완전성을 보였다.

정확성 측면에서 ChatGPT는 〈그림 6〉과 같이 주류 전반에서 8개 요소(Title, Creator, Publisher, Date, Format, Identifier, Language, Rights)를 우수한 성능으로 더블린 코어를 생성했다. 나머지 4개 요소의 경우, Description은 0.58이고 Contributor는 0.33이었으며 특히 Subject는 0.12, Type은 0.07로 매우 낮았다. 성능이 낮은



〈그림 5〉 DDC 주류별 더블린 코어 요소별 완전성



〈그림 6〉 DDC 주류별 더블린 코어 요소별 정확성

3개 요소 중에서 사회과학(300), 기술과학(600), 문학(800)은 다른 분야에 비해 더 낮은 정확성을 보였다.

실제 생성된 더블린 코어 내용을 살펴보면, Contributor를 거의 생성하지 않은 사회과학(300) 주류의 도서는 대부분이 국내 저자 작품이고 일부 해외 저자의 작품을 번역한 경우라고 하더라도 ChatGPT가 기여자(Contributor) 대신 번역자라고 구분한 것이 원인이다. 정확성 측면에서 Publisher는 완전성은 높았지만 생성된 내용이 출판사 대신 펴낸이 또는 발행인, 즉 인물의 이름이 선택되어 정확성이 다소 낮아진 모습이었다.

더불어 기술과학(600) 주류의 도서 역시 완전성 측면에서 Contributor를 절반 정도 생성하지 않았는데, 마찬가지로 이 주류의 도서는 대부분 국내 저자 작품이며, 일부 해외 저자의 작품을 번역한 경우라고 하더라도 ChatGPT가 기여자(Contributor) 대신 번역자라고 구분한 것이 원인이었다. 따라서 Contributor의 품질이 낮은 원인은 데이터 자체에서 메타데이터 요소용 서지 기술 내용의 누락, ChatGPT의 책임표시사항 추출 오류, ChatGPT가 사용하는

영어 위주의 학습데이터에 있다고 사료되었다. 이외에도 Rights도 완전성과 정확성이 다소 낮았는데, 이 주류의 도서는 다른 주류의 도서에 비해 판권기에 저작권 정보가 누락된 경우가 잦은 것이 원인으로 추측된다.

6. 종합 논의

이상의 연구 결과를 종합하면, ChatGPT는 완전성과 정확성 측면에서 요소 Title, Creator, Publisher, Date, Format, Identifier, Language, Rights를 매우 우수하게 생성할 수 있었다. 반면 Subject, Description, Contributor, Type에 대해서는 완전성 또는 정확성 문제를 보였다. 먼저 Subject, Description, Contributor 요소를 잘 만들지 못했으며 설사 만들어 내더라도 정확성이 매우 낮았다. 그다음 Type은 완전성 측면에서 매우 우수한 성능으로 생성되는 모습을 보였으나, 정확성은 매우 낮았다. 그럼에도 불구하고 Description 요소와 사례와 같이, 서술적 요소가 제공되지 않은 데이터로부터도 필요한 데이터를 유추하는 ChatGPT의 자체적 특

정이 드러났으며 그 작성하는 능력은 놀라운 수준이었다. <표 9>는 이러한 ChatGPT의 성능을 완전성과 정확성 수준에서 4가지로 구분하여 보여주는 사례이다.

위에서 언급한 요소들의 성능 문제의 원인은, 본 연구의 실험적 제약과 GPT 학습데이터 내 영어 데이터의 편중이 원인으로 의심되었다. 먼저 Subject는 국립중앙도서관의 주제

<표 9> ChatGPT가 생성한 더블린 코어를 요소별 성능으로 구분한 사례

a. 완전성 우수, 정확성 우수	b. 완전성 우수, 정확성 미흡
<p>한국어로 작성한 더블린 코어는 다음과 같습니다.</p> <ol style="list-style-type: none"> 1. 표제: 소크라테스 익스프레스 2. 창작자: 에릭 와이너 3. 주제: 철학 4. 설명: 철학이 일상 생활에 어떻게 적용될 수 있는지에 대한 내용 5. 발행자: 어크로스 출판그룹(주) 6. 기여자: 김하현 7. 날짜: 2021년 4월 28일(초판 1쇄), 2021년 7월 26일(초판 10쇄) 8. 유형: 책 9. 형식: 인쇄물 10. 식별자: ISBN 979-11-90030-92-2 11. 출처: 없음 12. 언어: 한국어 13. 관련 자원: 없음 14. 수록 범위: 철학과 일상 생활 사이의 연결 15. 이용 조건: 무단 전재와 무단 복제 금지, 서면 동의 필요 	<p>제목: 약간의 거리를 둔다 창작자: 소노 아야코 내용주제: 상처마저 거름이 되는 삶의 패러독스, 세상의 잣대로 나의 행복을 재단하지 마라! 설명: 소노 아야코의 에세이 발행자: 김현정 유형: 책 형식: 인쇄된 책 식별자: ISBN 979-11-86274-15-6 출처: 책읽는고양이 언어: 한국어 수록범위: 상처마저 거름이 되는 삶의 패러독스 이용조건: 판권 소유, 재배포 및 출판 금지</p>
c. 완전성 보통, 정확성 보통	d. 완전성 미흡, 정확성 미흡
<p>제목: 다 괜찮아요, 천국이 말했다 창작자: 미치 엘봄 옮긴이: 공경희 출판사: 살림출판사 출판일: 2020년 6월 22일 유형: 소설 형식: 인쇄된 책 식별자: ISBN 978-89-522-4197-9 언어: 한국어 관련자원: - 수록범위: - 이용조건: -</p>	<p>제목: 좋아하는 것을 돈으로 바꾸는 법</p> <ol style="list-style-type: none"> 1. 표제 (Title) 2. 창작자 (Creator): 멘탈리스트 다이코 3. 주제 (Subject) 4. 설명 (Description) 5. 발행자 (Publisher): (주)동양북스 6. 기여자 (Contributor): 김해용 7. 날짜 (Date): 2017년 9월 15일 8. 유형 (Type) 9. 형식 (Format) 10. 식별자 (Identifier): ISBN 979-11-5768-287-4 11. 출처 (Source) 12. 언어 (Language) 13. 관련자원 (Relation) 14. 수록범위 (Coverage) 15. 이용조건 (Rights)

명표목이 반영되어야 하는 요소이고 Type은 문서의 장르를 지정하는 통제어가 필요한 요소이다. 그리고 Contributor는 번역가, 편집자, 삽화가와 같이 도서를 발행하는 데 크게 이바지한 사람이나 기관이 언급되어야 하는 요소지만 ChatGPT에게 입력된 국립중앙도서관의 요소 설명(〈표 1〉 참조)은 다소 추상적이어서, ChatGPT가 책임표시사항을 구분하는데 무리가 따르는 것으로 생각되었기에 Subject, Type, Contributor의 성능이 낮은 원인은 본 연구의 실험환경에 의한 제약으로 파악되었다. 또한 Description은 도서 자체에 대한 설명이나 그 내용에 대한 설명을 포함해야 하는데, 이 요소의 완전성이 낮은 도서를 확인하니 국내 저자 또는 비영어권 저자가 쓴 것이 대부분이었다. 따라서 Description 요소의 성능이 낮은 이유는 GPT3 언어모델에서 비영어권 학습데이터 비중이 작기 때문으로 추측된다. 따라서 이러한 제약 사항을 개선하면 더블린 코어 요소별 성능은 더욱더 향상될 것으로 예상된다.

ChatGPT는 매우 강력한 도구임은 분명하다. 그러나 연구자의 의도대로 효과성을 극대화하려면 먼저 ChatGPT의 기술적인 특성을 이해하는 것이 선행되어야 한다.

첫 번째, 프롬프트 엔지니어링 개념을 인지해야 한다. 예를 들어 낯선 이와의 대화 도중 갑자기 “파이썬 프로그래밍 언어를 배우려면 어떻게 해야 하는가?”라는 질문을 받았을 때, 상대방에 대한 배경 정보가 충분치 않으면 정확하게 답변할 수 없다. 인간과 인간 사이의 인터페이스가 대화인 것처럼 인간과 인공지능 프로그램

의 인터페이스를 프롬프트라고 할 때, GPT3와 같은 대규모 언어모델이 인간의 요구를 이해하고 효과적으로 답변할 수 있도록 하려면 인간이 먼저 명확한 문장을 작성하고 답변에 필요한 정보를 제공해야 한다. 이처럼 인간이 프롬프트에 쓸 문장을 사려 깊게 작성하고 선정하는 과정을 프롬프트 엔지니어링이라고 한다(Rombach, Blattmann, & Ommer, 2022; Zhou et al., 2022).

두 번째, ChatGPT가 검색엔진이 아니다. 이는 2021년 학습이 완료된 GPT3 언어모델을 기반으로 한 챗봇 서비스일 뿐이므로 그 특성을 이해할 필요가 있다. 예를 들어, 입력데이터 중 『빨강 머리 여인』(ISBN 978-89-374-3785-4)은 튀르키예의 소설가 오르한 파묵의 소설이다. ChatGPT에게 이 도서에 대하여 알려달라는 요청을 한국어나 영어로 하면 루시 모드 몽고메리 『빨간 머리 앤』이나 아서 코난 도일의 셜록 홈스의 모험 시리즈 중 『빨간 머리 연맹』과 관련된 것으로 인식한다. 하지만 튀르키예어로 “Bana Orhan Ferit Pamuk’tan Kırmızı Saçlı Kadın’ı anlat.”⁷⁾라고 물으면 올바른 대답을 얻을 수 있다. GPT3 계열 모델 내에서 튀르키예어(tr) 비중⁸⁾은 한국어보다 높다.

세 번째, ChatGPT가 훌륭한 도구라면 사서는 도구를 뛰어넘는 존재여야 할 것이다. 예를 들어 Description의 경우 ChatGPT가 놀라운 정도의 문장 작성 능력을 보여주지만, 그 내용이 참인지 거짓인지는 사서가 판단해야 한다. 다음 두 문장은 본 연구 과정에서 김호연의 『불편한 편의점』(ISBN 979-11-6157-118-8)과 파울로 코엘료의 『연금술사』(ISBN 89-8281-447-7)에

7) 번역: Orhan Ferit Pamuk의 The Red Haired Woman에 대해 알려주세요.

8) 단어 수 기준 21위이며 비율은 0.05944%임. 한국어는 단어 수 기준 29위이며 비율은 0.01697%임.

대해 ChatGPT가 더블린 코어의 Description 요소값으로 생성한 것이다. 이 문장을 표지, 표제지, 판권기호만 판단했을 때 적합한 Description으로 판단되나, 외부 정보원을 사용해서 내용의 정확성을 살펴보면 내용에 있어 오류가 있다는 사실을 알 수 있다.

- 현대 사회에서 편의점은 매우 중요한 역할을 하지만, 그 속에서 일어나는 인간관계는 종종 불편한 상황을 초래한다. 이 소설은 이러한 사실을 바탕으로 인간관계의 복잡성을 그려낸다. **(소설 내 편의점은 불편한 상황을 초래하는 공간이 아니라, 다양한 인간 군상의 접점이다.)**
- 연금술사인 산티아고가 자신의 꿈을 따라 스페인에서 모로코, 이집트를 거쳐 최종 목적지인 상황의 동산으로 향하면서 겪는 여정과 그 과정에서 배우는 인생의 지혜를 다룬 소설. **(소설 내 산티아고는 연금술사가 아니라 양치기이다.)**

7. 결 론

이 연구는 사서가 정보 자원을 대상으로 메타데이터 부여와 같은 정보 조직 및 지적 작업 활동을 수행하듯이 ChatGPT도 이러한 업무가 가능한지 분석하기 위하여, ChatGPT에게 90권의 도서(인문사회과학분야 30권, 과학기술분야 30권, 문학분야 30권)로부터 수집한 표지, 표제지, 판권기 데이터를 입력하고 더블린 코어를 작성하도록 명령하여 결과를 얻었으며 이를 완전성과 정확성 척도로 평가하였다.

첫째, 전체 데이터에 대한 ChatGPT의 성능

에서 완전성은 0.87이며, 정확성은 0.71로 나타나, 전반적으로 양측 수준 모두 준수한 성능을 보였지만 정확성이 완전성보다 다소 낮은 성능으로, ChatGPT가 생성에 있어 더블린코어에 알맞게 추출은 잘하고 있으나, 추출한 실제의 결과에 대하여 상대적으로 불확실한 것으로 확인하였다.

둘째, ChatGPT는 입력한 도서 정보에서 기술 요소가 명확히 나타난 요소인 Title, Creator, Publisher, Date, Identifier, Language에 대하여 완전성과 정확성 모두에서 우수한 생성 능력을 보여주었다. 한편, 사서의 지적 사고에 의한 창출 활동이 요구되는 Subject나 Description 요소는 정확성에서 성능이 낮았지만, ChatGPT 자체의 생성 역량을 확인할 수 있는 요소인 것으로 확인하였다.

셋째, ChatGPT는 사회과학(300)이나 기술과학(600)에 해당하는 데이터에 대하여 전체 평균보다 다소 낮은 완전성과 정확성을 보였는데, 특히, Contributor의 품질 문제는 판권기 내 필요 데이터 누락, 책임표시사항 구분 오류, ChatGPT가 사용하는 학습데이터의 영어 편중이 원인으로 사료되었다. 따라서 ChatGPT 생성 성능을 높이기 위해서는 ChatGPT에게 추가 정보를 제공하는 방법을 고려할 수 있다. Subject의 경우 국립중앙도서관의 주제명 전거를 제시하고, Contributor는 편집자와 삽화가를 분류하는 예제를 입력할 수 있을 것이고, Type은 소설이나 수필 등 도서관에서 입력하는 장르 분류를 해야 할 것으로 예상된다. 단, GPT는 기술적으로 전통적 개념의 파인튜닝이 불가능하므로 외부 정보원을 참조해야 할 Source와 Relation과 Coverage는 성능 향상이 제한될 것

이다.

본 연구의 한계를 바탕으로 한 후속 연구의 제언은 다음과 같다. 첫째, ChatGPT의 제한적인 파인튜닝이 메타데이터 작성에 어느 정도의 긍정적인 영향을 미치는 정도를 파악하는 연구가 필요하다. 둘째, 영어 위주의 다국어 언어 학습 모델이 영어권 원작 번역서, 비영어권 원작 번역서, 한국어 원작의 메타데이터 작성에 미치는 영향을 파악하는 연구가 진행될 수 있다. 셋째, 더블린 코어 이외에도 실제 우리나라 도서관에서 많이 사용되는 MODS와 KOMARC

에 대해서도 자동 작성 능력을 평가할 필요가 있다.

본 연구는 자연어 인공지능 서비스의 특이점으로 호평받는 ChatGPT를 사용해서 도서 메타데이터의 자동 작성 가능성을 탐구했다는 데 의의가 있다. 인공지능은 도구이며, 도구는 사용자의 정확한 의도가 드러날 때 그 효과성이 극대화될 수 있다. 따라서 본 연구 결과를 통해 학계 연구자들이 기존의 문헌정보학 연구를 인공지능과 접목해 시너지 효과를 얻는 데에 이바지할 것으로 기대한다.

참 고 문 헌

- 노대원 (2023). 소설 쓰는 로봇: ChatGPT와 AI 생성 문학. 한국문예비평연구, 77, 125-160.
- 송학준, 송형용, 이지은 (2023). ChatGPT와 관광산업의 미래에 관한 연구. 호텔리조트연구, 22(1), 115-128.
- 양기철, 박정란 (2018). Automatic extraction of metadata information for library collections. The International Journal of Advanced Culture Technology, 6(2), 117-122.
<https://doi.org/10.17703/IJACT.2018.6.2.117>
- 용성중, 박효경, 유연희, 문일영 (2021). 영상 콘텐츠의 오디오 분석을 통한 메타데이터 자동 생성 방법. 한국향행학회논문지, 25(6), 557-561.
- 이경호 (2011). 정보학 (개정 3판). 대구: 인쇄마당.
- 이명규 (2010). 우리나라의 도서 판권기 기술서지 요소 고찰. 한국도서관·정보학회지, 41(1), 211-231.
<https://doi.org/10.16981/kliss.41.1.201003.211>
- 이용구, 김병규 (2011). 학술지 기사에 대한 메타데이터 품질의 계량화 방법에 관한 연구. 정보관리학회지, 28(1), 309-326. <https://doi.org/10.3743/KOSIM.2011.28.1.309>
- 이윤희, 김창식, 안현철 (2023). ChatGPT에 관한 연구: 뉴스 빅데이터 서비스와 ChatGPT 활용 사례를 중심으로. (사)디지털산업정보학회 논문지, 19(1), 139-151.
- 이치주, 이성숙, 김상규, 최성환, 국민상 (2000). 더블린 코어와 MARC. 도서관문화, 41(6), 4-34.
- 정종진, 김경원, 김구환 (2020). 데이터셋 검색 지원을 위한 메타데이터 자동 추출에 관한 연구. 한국통

- 신학회 학술대회논문집, 867-868.
- 정한민, 박정훈 (2023). ChatGPT를 이용한 문헌 작성 설계 및 이슈. 한국지식정보기술학회 논문지, 18(1), 31-40. <https://doi.org/10.34163/jkits.2023.18.1.004>
- Armengol-Estapé, J., Bonet, O. G., & Melero, M. (2021). On the Multilingual Capabilities of Very Large-Scale English Language Models. arXiv e-prints. <https://doi.org/10.48550/arXiv.2108.13349>
- Chapman, A. & Massey, O. (2002). A catalogue quality audit tool. Library management, 23(6/7), 314-324. <https://doi.org/10.1108/01435120210432282>
- Cox, C. & Tzoc, E. (2023). ChatGPT: Implications for academic libraries. College & Research Libraries News, 84(3), 99. <https://doi.org/10.5860/crln.84.3.99>
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint. <https://doi.org/10.48550/arXiv.1810.04805>
- Guinchard, C. (2002). Dublin Core use in libraries: a survey. OCLC Systems & Services: International digital library perspectives, 18(1), 40-50. <https://doi.org/10.1108/10650750210418190>
- Han, H., Giles, C. L., Manavoglu, E., Zha, H., Zhang, Z., & Fox, E. A. (2003). Automatic document metadata extraction using support vector machines. In 2003 Joint Conference on Digital Libraries, 37-48. <https://doi.org/10.1109/JCDL.2003.1204842>
- Huang, J., Shao, H., Chang, K. C. C., Xiong, J., & Hwu, W. M. (2022). Understanding jargon: Combining extraction and generation for definition modeling. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, 3994-4004.
- Irvin, K. M. (2003). Comparing information retrieval effectiveness of different metadata generation methods. Master's thesis, University of North Carolina at Chapel Hill, United States. <https://doi.org/10.17615/grff-0v98>
- James, R. & Weiss, A. (2012). An assessment of Google Books' metadata. Journal of Library Metadata, 12(1), 15-22. <https://doi.org/10.1080/19386389.2012.652566>
- Johnson, D., Goodman, R., Patrinely, J., Stone, C., Zimmerman, E., Donald, R., Chang, S., Berkowitz, S., Finn, A., Jahangir, E., Scoville, E., Reese, T., Friedman, D., Bastarache, J., Heijden, Y., Wright, J., Carter, N., Alexander, M., Choe, J., Chastain, C., Zic, J., Horst, S., Turker, I., Agarwal, R., Osmundson, E., Idrees, K., Kiernan, C., Padmanabhan, C., Bailey, C., Schlegel, C., Chambless, L., Gibson, M., Osterman, T., & Wheless, L. (2023). Assessing the accuracy and reliability of AI-generated medical responses: an evaluation of the Chat-gpt model. <https://doi.org/10.21203/rs.3.rs-2566942/v1>

- Kirtania, D. K. & Patra, S. K. (2023). OpenAI ChatGPT Generated Content and Similarity Index: A study of selected terms from the Library & Information Science (LIS). Qeios. <https://doi.org/10.32388/FO1CP6.3>
- Liu, V. & Chilton, L. B. (2022). Design guidelines for prompt engineering text-to-image generative models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 384, 1-23. <https://doi.org/10.1145/3491102.3501825>
- Lund, B. D. & Wang, T. (2023). Chatting about ChatGPT: how may AI and GPT impact academia and libraries?. *Library Hi Tech News*. <https://doi.org/10.1108/LHTN-01-2023-0009>
- Margaritopoulos, M., Margaritopoulos, T., Mavridis, I., & Manitsaris, A. (2012). Quantifying and measuring metadata completeness. *Journal of the American Society for Information Science and Technology*, 63(4), 724-737. <https://doi.org/10.1002/asi.21706>
- Moradi, M., Blagec, K., Haberl, F., & Samwald, M. (2021). Gpt-3 models are poor few-shot learners in the biomedical domain. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2109.02555>
- Ochoa, X. & Duval, E. (2009). Automatic evaluation of metadata quality in digital repositories. *International journal on digital libraries*, 10, 67-91. <https://doi.org/10.1007/s00799-009-0054-4>
- Ojokoh, B. A., Adewale, O. S., & Falaki, S. O. (2009). Automated document metadata extraction. *Journal of Information Science*, 35(5), 563-570. <https://doi.org/10.1177/0165551509105195>
- Park, J. R. (2009). Metadata quality in digital repositories: A survey of the current state of the art. *Cataloging & classification quarterly*, 47(3-4), 213-228. <https://doi.org/10.1080/01639370902737240>
- Qu, Y., Liu, P., Song, W., Liu, L., & Cheng, M. (2020). A text generation and prediction system: pre-training on new corpora using BERT and GPT-2. In *2020 IEEE 10th international conference on electronics information and emergency communication (ICEIEC)*, 323-326. IEEE. <https://doi.org/10.1109/ICEIEC49280.2020.9152352>
- Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training.
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, 1(8), 9.
- Rombach, R., Blattmann, A., & Ommer, B. (2022). Text-guided synthesis of artistic images with retrieval-augmented diffusion models. *arXiv preprint*.

- <https://doi.org/10.48550/arXiv.2207.13038>
- Sokvitne, L. (2000). An evaluation of the effectiveness of current Dublin Core metadata for retrieval. In VALA conference.
- Underwood, W. (2020). Automatic Extraction of Dublin Core Metadata from Presidential E-records. In 2020 IEEE International Conference on Big Data (Big Data), 1931-1938.
<https://doi.org/10.1109/BigData50022.2020.9377943>
- Valls-Vargas, J. (2013). Narrative extraction, processing and generation for interactive fiction and computer games. In Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, 9(6), 37-40. <https://doi.org/10.1609/aiide.v9i6.12600>
- White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). A prompt pattern catalog to enhance prompt engineering with chatgpt. arXiv preprint. <https://doi.org/10.48550/arXiv.2302.11382>
- Zavalina, O. L. & Burke, M. (2021). Assessing skill building in metadata instruction: quality evaluation of dublin core metadata records created by graduate students. *Journal of Education for Library and Information Science*, 62(4), 423-442.
<https://doi.org/10.3138/jelis.62-4-2020-0083>
- Zhang, J. & Dimitroff, A. (2004). Internet search engines' response to metadata Dublin Core implementation. *Journal of Information Science*, 30(4), 310-320.
<https://doi.org/10.1177/0165551504045851>
- Zhang, J. & Dimitroff, A. (2005a). The impact of webpage content characteristics on webpage visibility in search engine results (Part I). *Information Processing & Management*, 41(3), 665-690. <https://doi.org/10.1016/j.ipm.2003.12.001>
- Zhang, J. & Dimitroff, A. (2005b). The impact of metadata implementation on webpage visibility in search engine results (Part II). *Information processing & management*, 41(3), 691-715. <https://doi.org/10.1016/j.ipm.2003.12.002>
- Zhou, Y., Muresanu, A. I., Han, Z., Paster, K., Pitis, S., Chan, H., & Ba, J. (2022). Large language models are human-level prompt engineers. arXiv preprint.
<https://doi.org/10.48550/arXiv.2211.01910>

• 국문 참고문헌에 대한 영문 표기

(English translation of references written in Korean)

Jung, Hanmin & Park, Junghun (2023). Design and issues of writing literatures using ChatGPT.

- Journal of Knowledge Information Technology and Systems, 18(1), 31-40.
<https://doi.org/10.34163/jkits.2023.18.1.004>
- Jung, Jong Jin, Kim, Kyung Won, & Kim, Gu Hwan (2020). A Study on automatic metadata extraction to support dataset search. Proceedings of KICS Summer Conference 2020, 867-868.
- Lee, Chi-Ju, Lee, Sung-Sook, Kim, Sang-Gyu, Choi, Sung-Hwan, & Kook, Min-Sang (2000). Dublin core and MARC. KLA Buttetin, 41(6), 4-34.
- Lee, Kyungho (2013). Information Science (3rd ed.). Daegu: Inswaemadang.
- Lee, Myounggyu (2010). A study on the description elements of the book colophon in Korea. Journal of Korean Library and Information Science Society, 41(1), 211-231.
<https://doi.org/10.16981/kliss.41.1.201003.211>
- Lee, Yong-Gu & Kim, Byungkyu (2011). A study on quantitative measurement of metadata quality for journal articles. Journal of the Korean Society for Information Management, 28(1), 309-326. <https://doi.org/10.3743/KOSIM.2011.28.1.309>
- Lee, Yunhee, Kim, Changsik, & Ahn, Hyunchul (2023). A study on the ChatGPT: focused on the news big data service and ChatGPT use cases. Journal of the Korea Society of Digital Industry and Information Management, 19(1), 139-151.
- Noh, Dae-won (2023). Fiction-writing robot: ChatGPT and AI-generated literature. Journal of Korean Literary Criticism, 77, 125-160.
- Song, Hak Jun, Song, Hyoung-yong, & Lee, JiEun (2023). A study on the future of tourism industry and ChatGPT. Journal of Hotel & Resort, 22(1), 115-128.
- Yang, Gi-Chul & Park, Jeong-Ran (2018). Automatic extraction of metadata information for library collections. The International Journal of Advanced Culture Technology, 6(2), 117-122.
<https://doi.org/10.17703/IJACT.2018.6.2.117>
- Yong, Sung-Jung, Park, Hyo-Gyeong, You, Yeon-Hwi, & Moon, Il-Young (2021). Method of automatically generating metadata through audio analysis of video content. Journal of Advanced Navigation Technology, 25(6), 557-561.