

ISSN 2765-3005

국립중앙도서관 이슈페이퍼

본문검색 서비스 현황 및 발전방안

Vol. 19

2024. 2.

ISSUE PAPER

NCK



문화체육관광부
국립중앙도서관

본문검색 서비스 현황 및 발전방안

김재균 주무관 | 국립중앙도서관 정보기술기반과

I. 들어가며	02
II. 서비스 현황 및 검색기술	02
1. 서비스 현황	02
2. 검색 적용 기술	04
III. 본문검색 서비스 사례 및 저작권	07
1. 서비스 사례	07
2. 저작권	09
IV. 생성형 인공지능(AI) 기반 검색과 본문검색 발전방안	11
1. 생성형 AI 검색 서비스 사례	11
2. 본문검색 서비스 발전방안	13
참고문헌	15

주요 키워드

검색 서비스, 본문검색, 저작권, 인공지능, AI



'공공누리' 출처표시-상업적 이용금지-변경금지 조건에 따라 이용할 수 있습니다. 단, 표지, 일러스트레이션 및 사진은 제외됩니다.

국립중앙도서관 이슈페이퍼 제19호

발행일 2024년 2월 29일

발행처 국립중앙도서관

발행인 국립중앙도서관장

주소 서울특별시 서초구 반포대로 201

전화 02-590-0799

팩스 02-590-0546

누리집 <https://nl.go.kr>

ISSN 2765-3005

• 본지에 실린 글의 내용은 집필자의 개인적인 견해이며, 국립중앙도서관의 공식적인 의견과 다를 수 있습니다.

• 본지의 저작권은 국립중앙도서관에 있으며, 사전 허락 없이 무단으로 복제·변경·배포할 수 없습니다.

본문검색 서비스 현황 및 발전방안

김재균 주무관

(국립중앙도서관 정보기술기반과)

요약

국립중앙도서관 본문검색 서비스는 이용자가 도서관 소장자료와 학술논문의 본문 내용을 검색할 수 있는 서비스이다. 제목이나 저자 등의 정보를 알지 못해도 검색한 키워드와 일치하는 부분을 본문 내용에서 찾아볼 수 있다. 시범적으로 2011년부터 2015년까지 발행된 일반도서자료 8만 8천 권과 학술논문자료 1백만 건의 본문 텍스트를 대상으로 색인을 구축하고, 2022년부터 본문검색 서비스를 제공하고 있다. 본문검색 서비스는 국내외 여러 곳에서 유사한 서비스를 제공하는데 대표적으로 구글 도서(Google Books), 오픈 라이브러리(Open Library) 등이 있다. 본문검색 서비스는 결과화면에서 본문 내용을 미리 확인할 수 있어 저작권 문제가 발생하는데 저작권 논쟁은 2006년부터 2010년 사이에 활발하게 논의되었고, 합의안을 거쳐 일정 비율 이하의 내용만 제한하여 서비스하는 것으로 일단락되었다.

본문검색 서비스에 생성형 인공지능(AI) 기술을 결합한 서비스로는 본문 내용으로 구축한 색인을 기반으로 데이터를 학습시킨 후 단순 키워드 검색이 아닌 질문의 의미를 분석하고 이를 본문 내용과 연결하여 이용자가 정확한 검색어를 알지 못하더라도 원하는 검색 결과를 제공하는 AI 도서 검색 서비스를 고려해 볼 수 있다.

주요 키워드 검색 서비스, 본문검색, 저작권, 인공지능, AI

I. 들어가며

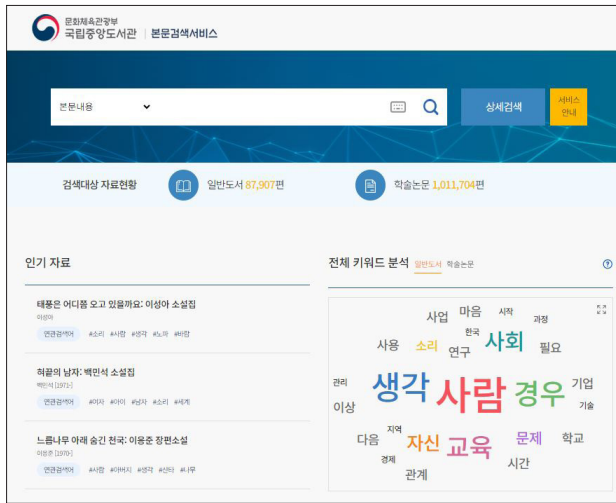
본문검색은 도서의 본문 내용을 검색하여 찾아볼 수 있는 서비스이다. 전통적인 도서 검색에서는 제한적으로 정해진 항목(필드) 내에서만 검색을 수행하던 것과 달리 이용자가 본문 내용을 검색어로 활용할 수 있는 이용자 주도형 정보서비스라고 할 수 있다. 본문검색은 서로 다른 용어로 사용되는데 우리나라에서는 ‘본문검색 서비스’로, 서비스를 처음 시작한 아마존에서는 ‘Search inside the book’, 구글은 ‘Google Book Search’ 등으로 사용하고 있다. 이 글에서는 본문검색 서비스 현황과 기술, 사례를 알아보고 이를 도서관 누리집에 어떻게 활용하여 발전시켜 나갈지 논의해 본다.

II. 서비스 현황 및 검색기술

1. 서비스 현황

국립중앙도서관에서는 2021년 본문검색 시범 서비스 구축 사업을 진행하여 자료 110만여 건에 대해 본문검색 서비스(<https://www.nl.go.kr/NLFP>)¹를 제공하고 있다. 일반도서 8만 8천 건, 학술논문 1백만 건에 대해 시범적으로 본문 텍스트를 추출하여 검색 시스템을 구축하였다. 이용자가 입력한 키워드를 본문 내용에서 찾아 결과로 제공한다. 문장 또는 키워드로 본문 내용을 검색하여 자료유형/저작권유형별로 검색된 자료 목록을 책단위로 확인할 수 있다. 또한 미리보기 기능으로 페이지 정보와 매칭되는 본문 내용 일부를 보여주고, 원문보기를 누르면 전용 원문 뷰어로 전체 내용을 직접 확인할 수 있다. [그림 1]은 국립중앙도서관 본문검색 서비스 메인화면과 검색 결과화면이다.

¹ 현재는 국립중앙도서관 관내에서만 이용 가능.



[그림 1] 국립중앙도서관 본문검색 서비스

본문검색 서비스를 효율적으로 이용하도록 특화된 아키텍처를 적용하여 원문 PDF의 안정적인 수집 및 관리 시스템을 구축하였고, OCR PDF를 고품질로 서비스하고자 교정 관리 기능을 적용하였다(그림 2 참조). 또한 본문 내용을 효율적으로 추출·색인·검색하려고 검색 체계를 구축하였다.



[그림 2] 본문검색 교정·교열 화면

2. 검색 적용 기술

기존에 제공 중인 국립중앙도서관 누리집 검색방식은 검색어 키워드 완전일치 또는 부분일치 검색을 활용하여 검색 결과를 출력한다. 매칭되는 검색 결과를 보여주려고 검색엔진에서 미리 구축해 놓은 정보를 색인이라고 한다. 도서 색인구축은 일반적으로 표제, 저자, 발행자, 청구기호, 목차 등의 정보를 활용한다. 본문검색 서비스를 제공하려면 도서 본문 전체 텍스트를 대상으로 색인을 구축하는데, 원문구축 방식에 따라 텍스트를 그대로 추출할 수 있는 원문 PDF와 OCR 변환자료를 활용할 수 있다. OCR 변환자료는 내용에 따라 인식률에 차이가 있는데, 원본과 비교했을 때 90% 이상의 인식률은 되어야 원활한 서비스를 제공할 수 있다.

그리고 본문검색에 특화된 시스템을 설계하고자 클라우드 기반의 오픈 소프트웨어(SW) 검색엔진을 적용하였고, 형태소 기반 색인을 구축하여 검색 서비스를 구현하였다. OCR 추출 TXT, XML 선/후처리 공정을 활용한 수집/색인 최적화를 적용하였다.

가. 검색엔진

최적의 검색엔진을 적용하려 테스트 자료로 OCR PDF 18종을 가지고 4종 검색엔진의 색인속도, 색인용량, 검색속도, 검색품질을 기준으로 점수를 산출하였다. <표 1>은 검색엔진 4종의 SW별 벤치마킹 종합산정표이다.

<표 1> 검색엔진 SW별 벤치마킹 종합산정표

구분		코난테크놀로지 SEARCH5			바이브 MSIR5.0			아파치 Solr 8.2			일래스틱서치 7.1		
특징		국립중앙도서관 대표 누리집 통합검색에서 사용 중			(구)서지정보유통지원 시스템에서 사용			오픈 SW			오픈 SW		
문서수집 및 색인 모듈 여부		있음			있음			있음			없음		
이슈사항		검색속도가 다소 느림			검색속도가 다소 느림			오픈 SW로 기술지원 제약			오픈 SW로 기술지원 제약		
성능측정		측정 결과	점수	가중치 적용	측정 결과	점수	가중치 적용	측정 결과	점수	가중치 적용	측정 결과	점수	가중치 적용
색인속도	N-gram	31	5	0.5	51	3	0.3	55	3	0.3	59	3	0.3
	형태소	24	5	0.5	53	3	0.3	51	3	0.3	58	3	0.3

색인용량	N-gram	641M	2	0.4	25M	5	1	134M	4	0.8	87M	3	0.6
	형태소	408M	2	0.4	22M	5	1	21M	5	1	37M	5	1
검색속도	N-gram	6.7ms	3	0.9	5.7ms	3	0.9	4.0ms	5	1.5	3.9ms	5	1.5
	형태소	1.475ms	3	0.9	1.25ms	3	0.9	0.875ms	5	1.5	1.0ms	4	1.2
검색품질	N-gram	4건	4	0.8	4건	4	0.8	5건	5	1	3건	3	0.6
	형태소	6건	3	0.6	6건	3	0.6	26건	5	1	4건	2	0.4
가격		-	2	0.2	-	3	0.3	0	5	0.5	0	5	0.5
최종점수				5.2			6.1			7.9			6.4

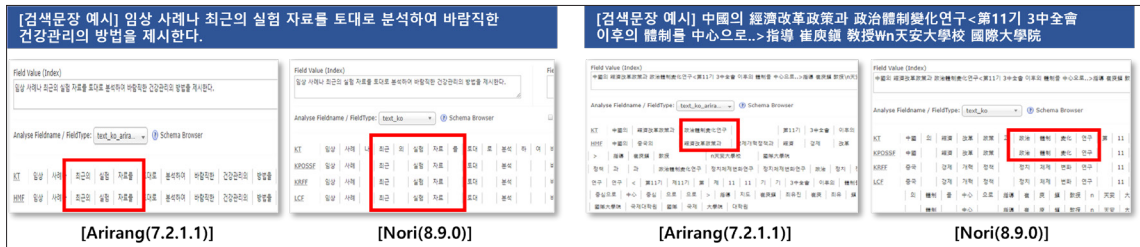
검색엔진 4종을 비교하였을 때 종합적으로 아파치 Solr가 우수한 점수를 얻어 도입하기로 결정하였다. 아파치 Solr는 루씬² 기반의 검색 기능을 제공하는 오픈소스 검색 플랫폼으로 실시간 인덱싱, 전문검색, 동적 클러스터링 등 사용자에게 다양한 기능을 제공하며 JSON, XML, CSV, HTTP Request 등 다양한 데이터 타입의 색인이 가능하다. 또한 검색한 데이터를 JSON, XML, CSV, 바이너리 타입으로 결과를 제공해 줄 수 있다. 전체 시스템 구조는 Solr에 대한 클러스터 구조를 확장성 있게 구성하고자 Solr 클라우드 모드를 통해 서버 4대를 하나의 논리 파티션으로 클러스터링하여 분산 처리하도록 구성하였다. 또한 고가용성을 보장하고자 특정 노드에 장애가 발생할 때 대체 노드로 장애를 극복하도록 구성하고 각 노드의 상태 체크 등의 기능을 적용하였다.

나. 형태소 분석기

형태소 분석기는 자연어 처리에서 활용되는 프로그램의 하나로, 입력된 텍스트를 명사, 동사, 형용사, 부사, 조사, 어미 등의 품사 및 형태소 정보가 붙어 있는 데이터로 분석해 주는 기능을 수행한다.

형태소 분석기는 성능이나 분석 기준 등이 조금씩 다르며, 다양한 오픈소스가 공개되어 있는데 그중에서 아리랑(Arirang)과 노리(Nori) 형태소 분석기를 선정하여 비교하였다. 형태소 분석, 검색어와 문서 분석, 검색어 띄어쓰기 등의 기능에 대해 비교 테스트를 진행하였다. 문장 내 형태소 분석결과에서 노리가 아리랑 대비 형태소 분석 및 한자 독음 처리 개수가 많이 출력되었다([그림 3] 참조).

² 아파치 재단에서 지원하는 자바 언어로 이루어진 정보 검색 라이브러리.



[그림 3] Arirang과 Nori 형태소 분석 결과 예시

노리 형태소 분석기의 경우 검색어의 형태소를 더 작은 단위로 분리하여 의미 있는 단어 조합으로 검색 시 문서에서 추출된 형태소의 개수가 많아 검색 재현율이 높은 장점을 보였다. 또한 띄어쓰기가 없어도 형태소를 최대한 정확하게 분리하는 부분에서 노리 형태소 분석기가 좀 더 나은 성능을 보였다. 이러한 비교 테스트에서 검색 결과 품질이 우위일 것으로 보이는 노리 형태소 분석기를 적용하기로 결정하였다.

다. 수집/관리 기능

본문검색 시스템에서는 자료 단위 또는 유형/연도/저작권 등의 조건으로 수집 설정을 관리한다. 수집 조건에 따라 원문자료 및 서지정보를 수집하고 수집한 원문의 본문 텍스트를 추출한다. 수집한 자료에 대하여 원문/교정·교열 동시 보기 기능과 작업 이력 조회, 임시저장, 작업완료 등의 관리가 가능하도록 구현하였다. 그리고 서비스 자료 설정 관리 기능으로 검색 대상 범위를 설정하도록 하였다.

라. 검색 기능

검색 수행 시 완전일치/부분일치/목차·초록 검색 등 다양한 검색 옵션으로 검색을 수행하고 정확도/최신순 등의 정렬, 유형/연도/소장처 등의 검색 필터를 이용해 검색 결과를 효율적으로 제공한다. 또한 검색 키워드 하이라이팅, 워드클라우드, 한자독음변환 등 다양한 사용자 편의 기능이 적용되었다. 책 단위/페이지 단위 검색으로 검색 결과 클러스터링 및 본문 텍스트 미리보기 등이 가능하고 원문 뷰어 연동으로 해당 원문을 확인할 수 있게 설계되었다.

마. 키워드 분석

본문검색 서비스에서는 수집된 OCR 및 TEXT PDF 자료의 텍스트를 형태소 분석으로 키워드 출현 빈도를 추출하여 서비스를 제공한다. 전체 키워드, 두 글자(bi-gram) 키워드, 네 글자(4-gram) 이상 키워드로 분류하여 국립중앙도서관 자료에 대한 빅데이터 기반으로 활용할 수 있다. 본문검색 서비스 메인화면에서는 두 글자 키워드에 대한 자료 유형별 주요 키워드를 워드 클라우드 형태로 한눈에 살펴보게 해놓았다. <표 2>에서는 두 글자 키워드의 출현 빈도 상위 10개를 보여준다.

<표 2> 두 글자(bi-gram) 키워드의 출현 빈도 상위 10위

순위	키워드	빈도수
1	사람	1,883,794
2	생각	1,172,343
3	경우	909,914
4	교육	839,805
5	사회	823,734
6	자신	813,439
7	문제	718,283
8	소리	690,958
9	연구	657,640
10	필요	636,579

Ⅲ. 본문검색 서비스 사례 및 저작권

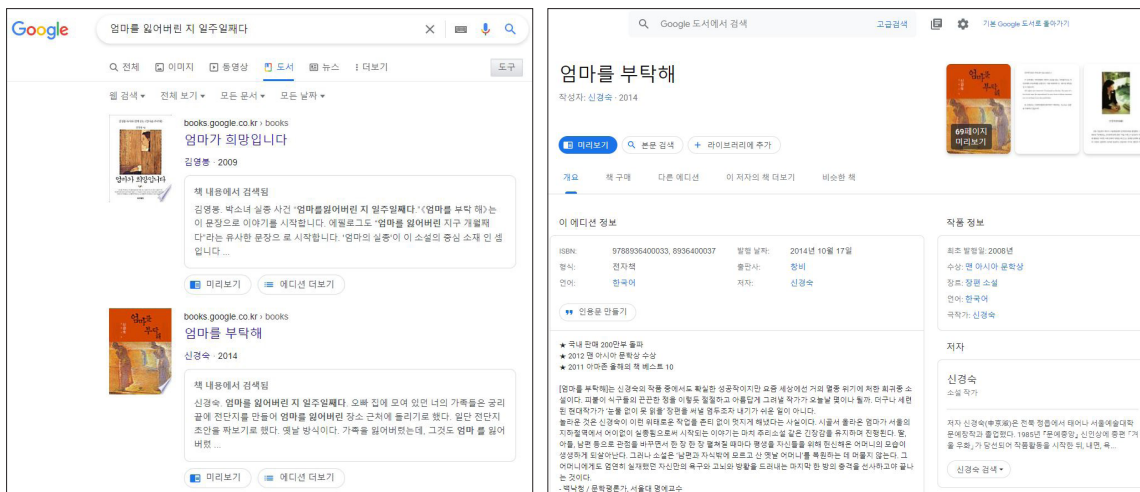
1. 서비스 사례

가. 구글 도서(Google Books)

2004년 12월 구글은 하버드, 스탠퍼드 등의 대학 및 뉴욕 공립도서관과 협약을 맺고 소장도서들을 스캔하여 데이터베이스를 구축하는 구글 도서관 프로젝트를 발표하였다. 이후 구글은 전 세계의 주요 도서관들과 협력해 도서와 학술자료 등을 디지털화하여 인터넷상에서 언제 어디서나 검색할 수 있는

디지털 도서관 구축을 추진하고 있다. 구글 도서(<https://books.google.co.kr>)는 구글 도서관 프로젝트로 구축된 전 세계 도서 수천만 권을 검색할 수 있는 서비스로 제목, 저자, 발행자, 발행일, ISBN 등 상세검색은 물론 본문 내용 검색과 검색 결과에 따른 미리보기, 원문보기 등을 제공하고 있다.

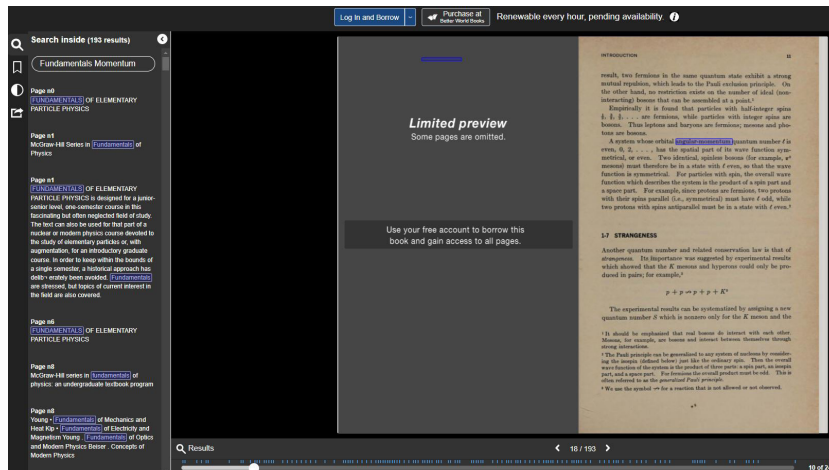
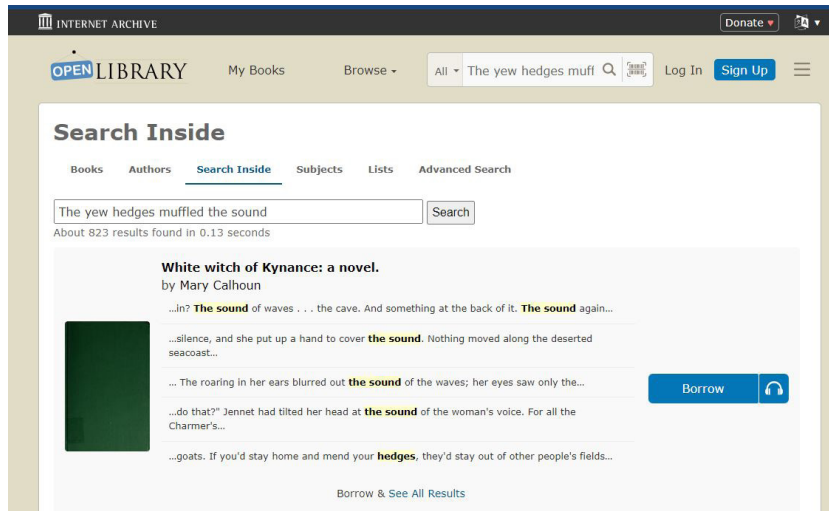
본문검색은 도서의 저작권에 따라 단계별 서비스를 제공한다. 우선 저작권이 없는 공개된 도서와 출판사가 허가한 도서는 무료로 원문 전체보기를 제공한다. 소유자가 확인된 저작권 보호 자료는 일반적으로 출판사에서 정한 비율만큼 미리보기 서비스를 제공하며, 복사·다운로드·인쇄가 제한된다. 소유자가 불분명하거나 소유자가 미리보기를 거부한 도서는 세 문장 이하의 발췌문을 제공하고, 디지털화되지 않은 도서는 미리보기 없이 일반적인 서지정보만 제공한다. [그림 4]는 Google Books 서비스 화면이다.



[그림 4] Google Books 서비스 화면

나. Open Library

Open Library는 미국의 비영리단체인 인터넷 아카이브(INTERNET ARCHIVE)에서 운영하는 서비스로 모든 출판 도서를 하나의 웹페이지에서 제공하는 온라인 프로젝트이다. Open Library 사이트(<https://openlibrary.org>)에서는 도서 본문검색 기능을 포함한 다양한 검색 기능을 제공하며, 140만 권 이상의 해외 도서 원문을 서비스하고 있다. 검색 옵션 중 Search Inside를 활용하여 본문검색이 가능하며, 원문이 있는 경우 상세보기로 들어가면 검색한 키워드와 일치하는 내용을 전부 확인할 수 있다. [그림 5]는 Open Library 본문검색 서비스 화면이다.



[그림 5] Open Library 본문검색 서비스 화면

2. 저작권

가. 국내

우리나라는 저작권법상 도서가 디지털 형태로 판매되면 디지털 복제가 금지되어 있다.³ 2006년 네이버와 다음에서 도서 본문검색 서비스를 시작하면서 출판계에서는 본문검색 서비스가 저작권 침해는 물론 도서 판매량 저하로 이어질 것을 우려하였다. 이에 2006년 11월 출판계와 교보문고, 다음, 네이버,

³ 저작권법 제31조(도서관에서의 복제 등) 제4항 도서관등은 제1항제2호의 규정에 따른 도서등의 복제 및 제2항과 제3항의 규정에 따른 도서등의 복제의 경우에 그 도서등이 디지털 형태로 판매되고 있는 때에는 그 도서등을 디지털 형태로 복제할 수 없다.

북토피아는 ‘도서 본문검색 서비스 기준안’을 마련하고 협약을 체결하였다. <표 3>은 도서 본문검색 서비스 기준안 중 서비스 범위를 발췌한 내용이다.

<표 3> 도서 본문검색 서비스 기준안

제3조(내용 및 범위)

① 본문 검색 서비스 범위

- 가. 본문 검색 서비스는 각 도서별로 30일간 누적해 전체 쪽수의 최대 5% 범위 내에서 제공할 수 있다. 단, 100쪽 미만 도서의 경우 5쪽까지 보여줄 수 있다.
- 나. 본문 검색을 서비스하고자 하는 업체는 가항의 범위 내에서, 출판사가 개별 도서에 대해 5% 미만의 검색 총량을 비율과 페이지 단위로 세분해서 지정할 수 있는 관리 페이지를 제공하여야 한다.

나. 미국

2006년 구글 도서관 프로젝트를 시작하면서 미국작가조합과 미국출판협회는 구글을 상대로 무분별한 도서의 디지털화 및 본문검색 서비스에 대한 저작권 침해 소송을 제기하였다. 구글이 제안한 것은 작가의 저작권 동의 없이도 전체 책을 이용하는 권리를 주는 것이었는데, 미리 동의를 구하고 디지털화하는 것이 아니라 우선 디지털화하여 서비스한 후 저작자가 이의를 제기하는 경우에 한해 서비스를 중지하는 것이기 때문에 저작권 문제가 커졌다. 논쟁의 핵심은 구글의 디지털 복제 및 도서 본문검색 서비스가 저작권법이 인정하는 공정이용에 해당하는지였다.

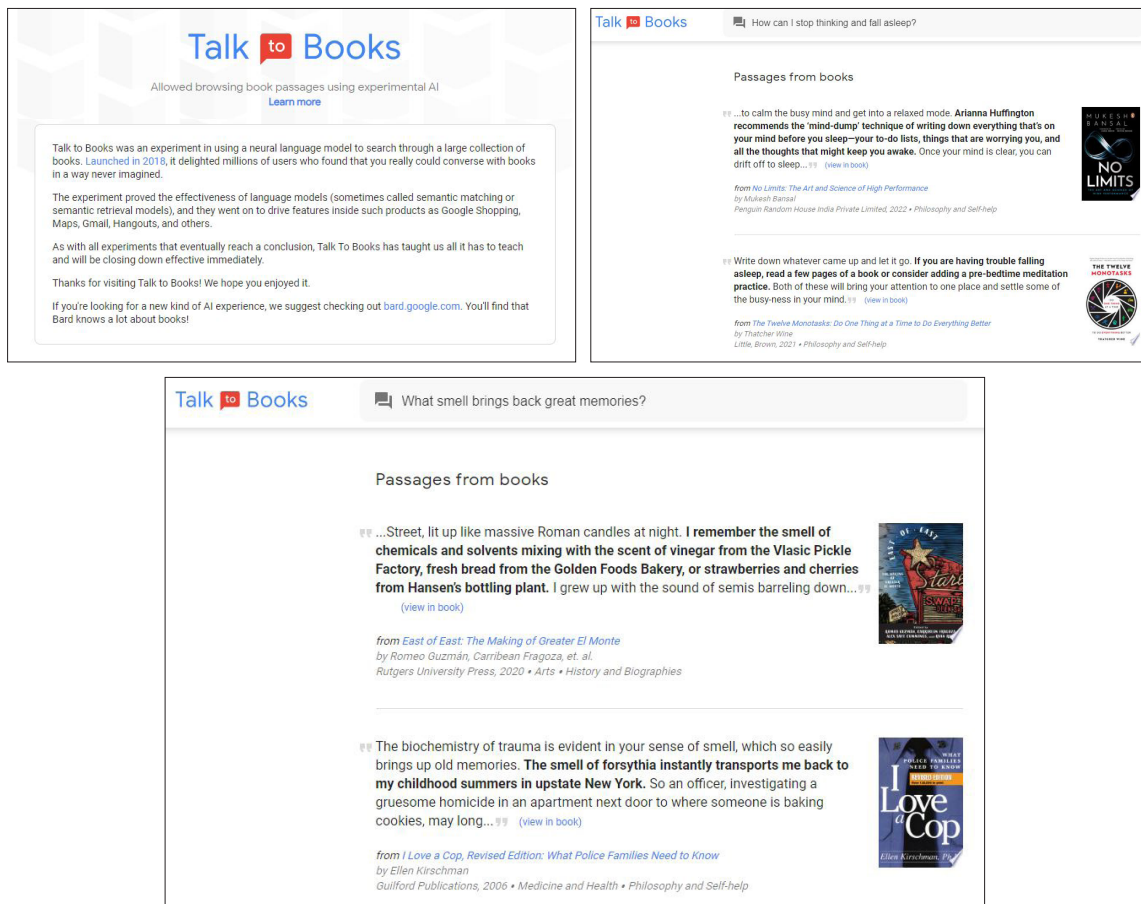
2016년 4월 미국 연방대법원은 10년 이상 지속된 저작권 소송에서 구글 도서관 프로젝트가 공정이용에 해당한다고 판결한 연방항소법원의 결정을 유지하고, 미국작가협회의 상고 신청을 기각하였다. 대법원은 구글이 책 수백만 권을 스캔한 것은 저작권법상 공정이용에 해당한다는 하급법원 판결을 심리하지 않기로 했다고 밝혔다. ‘공정이용’ 공방에서 핵심 쟁점은 이용목적, 원작의 성격, 이용 분량, 시장 피해 여부 등이었다. 구글은 공정이용 측면에서 자신들의 작업이 이용자들의 편의를 향상하는 것이라고 주장하였다. 항소심 재판부는 구글 도서가 목록을 이용해 책을 찾기 쉽게 해주기 때문에 오히려 원작품에 이익을 준다는 것이 법원의 판결이었다.

Ⅳ. 생성형 인공지능(AI) 기반 검색과 본문검색 발전방안

1. 생성형 AI 검색 서비스 사례

가. 구글 톡투북스(Talk to Books)

‘구글 톡투북스(Talk to Books)’ 서비스는 구글에서 제공하는 인공지능(AI) 기반 도서 검색 서비스로 일반적인 도서 검색과 달리 이용자가 대화 형식으로 질문을 던질 수 있다. 도서 내용이나 정보에 대한 질문을 던지면, 질문의 의미를 이해하여 사용자가 원하는 도서를 검색해 주는 서비스로 질문과 관련된 본문 내용을 책과 함께 미리보기 형태로 찾아준다. [그림 6]은 구글 톡투북스 서비스 화면이다.

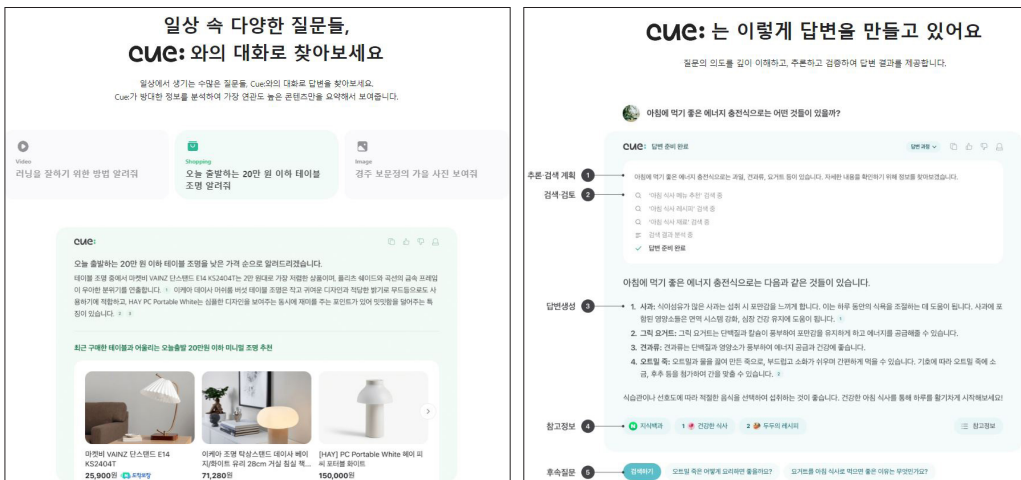


[그림 6] 구글 톡투북스(Google Talk to Books)

구글에서 실험적인 AI 서비스 모델로 개발하여 2018년부터 2023년 상반기까지 운영하였지만 현재는 운영이 중단되었다. 그 대신 구글에서 2023년 공개한 바드(Bard, 현 제미니(Gemini))에서 유사한 서비스를 활용할 수 있다.

나. 네이버 큐(Cue:)

Cue:는 네이버에서 개발 중인 AI 검색 서비스로 사용자가 대화 형식으로 질문을 던지면 질문의 의미를 이해하여 사람처럼 판단하고 관련 자료를 검색한 후 필요한 문서를 찾아 답변을 생성해 준다. 또한 이전의 질문과 답변을 기억하는 상태에서 대화 형식으로 계속 검색을 이어갈 수 있다. Cue: 서비스는 현재 일부 사용자를 대상으로 제한적인 서비스를 제공하고 있다. [그림 7]은 네이버 Cue: 서비스 안내 화면이다.



[그림 7] 네이버 Cue:

Cue:는 하나의 대형 언어 모델(LLM)을 사용하는 것이 아니라 모듈화된 LLM 플랫폼으로 설계하여 사용자의 검색 목적에 필요한 언어 모델만 작동시켜 효율성과 응답 속도를 올렸다. 사용자 질의가 들어오면 추론으로 검색 목적을 이해하고 검색하여 검색 결과 중 신뢰성 있는 결과를 선택해서 원하는 답변을 생성해 준다.

2. 본문검색 서비스 발전방안

국립중앙도서관은 저작권법⁴에 따라 자료 보존을 위한 수집 목적으로 디지털 복제를 인정받고 있으며, 디지털화 작업을 지속해 도서 원문 파일을 수집·보존하고 있다. 디지털화로 생성된 원문 파일에서 추출한 도서의 본문을 데이터 셋으로 구축하여 서지정보와 함께 생성형 AI와 결합함으로써 검색에 활용할 수 있다. 생성형 AI를 위한 학습데이터로 본문 데이터를 사용해도 되는지는 저작권법에 명확하게 나와 있지 않으며, 현재 입법 논의가 진행되고 있다. AI 학습에 대한 공정이용 인정 여부를 판단하고, 결과물이 기존 저작물의 저작권을 침해할 가능성이 있는지 확인하는 과정이 필요하다. 저작권법 제35조의5에 따르면 저작자의 정당한 이익을 부당하게 해치지 않으면 저작물을 이용할 수 있다(<표 4> 참조).

<표 4> 저작권법 제35조의5(저작물의 공정한 이용)

저작권법 제35조의5(저작물의 공정한 이용)

- ① 제23조부터 제35조의5까지, 제101조의3부터 제101조의5까지의 경우 외에 저작물의 일반적인 이용 방법과 충돌하지 아니하고 저작자의 정당한 이익을 부당하게 해치지 아니하는 경우에는 저작물을 이용할 수 있다.
- ② 저작물 이용 행위가 제1항에 해당하는지를 판단할 때에는 다음 각호의 사항 등을 고려하여야 한다.
 1. 이용의 목적 및 성격
 2. 저작물의 종류 및 용도
 3. 이용된 부분이 저작물 전체에서 차지하는 비중과 그 중요성
 4. 저작물의 이용이 그 저작물의 현재 시장 또는 가치나 잠재적인 시장 또는 가치에 미치는 영향

저작권 침해가 없을 수준의 적절한 도서 검색 서비스 모델을 만들어 데이터 셋을 학습하게 한 뒤 위의 사례에서 소개했듯이 ‘구글 톡투북스’에서 제공하는, 사용자가 정확한 키워드를 알지 못하더라도 질문을 만들어 입력하면 관련된 도서를 찾아주는 서비스를 구현해 볼 수 있다. 결과화면에서는 일정 비율 이상의 본문 내용을 볼 수 없게 제한하는 등 심도 있는 검토가 필요하다.

이러한 생성형 AI 기반 도서 검색 서비스로 이용자들은 정확한 제목이나 키워드를 알지 못한 상태에서 찾고자 하는 자료들을 쉽게 검색할 수 있고, 이용자가 필요하지만 잘 알려지지 않은 도서를 검색해 발견하도록 도울 수 있다. 출판사나 판매자로서도 도서 홍보 및 판매 연결에 긍정적으로 작용할 수 있다. 실제로 최근 버클리(Berkely)와 노스이스턴(Northeastern) 학자들의 연구(Nagaraj & Reimers, 2023)에 따르면, 구글의 도서 디지털화와 본문검색이 도서 판매량 증가에 영향을 주었다는 것을 알 수 있다.

⁴ 저작권법 제31조(도서관등에서의 복제 등) 제8항 「도서관법」 제22조에 따라 국립중앙도서관이 온라인 자료의 보존을 위하여 수집하는 경우에는 해당 자료를 복제할 수 있다.

그리고 AI 기술을 활용하여 도서 본문검색과 더불어 맞춤도서 추천 서비스 등 개별 이용자들이 원하는 방향의 검색 결과를 제공하도록 검색 서비스를 더욱 정교하게 개선한다면 이용자 편의성 제고와 디지털 정보격차 해소에 기여할 것이다.

참고문헌

네이버 Cue. <https://cue.search.naver.com>

신용우 (2023). 생성형 AI 관련 저작권 쟁점과 대응 방안 (문화정보 이슈리포트 2023-2호). 한국문화정보원.

위키피디아 Google Books. https://en.wikipedia.org/wiki/Google_Books/

이용준 (2010). 도서본문검색 서비스의 현황과 발전방안. 동서언론, 13, 37-60.

한국저작권위원회 (2016). 저작권 동향 (2016년 제8호).

Google Books. <https://books.google.co.kr>

Google Talk to Books. <https://books.google.com/talktobooks/>

Nagaraj, A. & Reimers, I. (2023). Digitization and the Market for Physical Works: Evidence from the Google Books Project. American Economic Journal: Economic Policy, 15(4), 428-458.

OPEN LIBRARY. <https://openlibrary.org>