

빅데이터, 인공지능을 활용한 도서관 장서 포화문제 해결 연구 - 대구 두류도서관을 사례로 -

| 요약

- 본 연구에서는 공공도서관의 장서 포화문제를 해결하기 위해 최근 4차 산업혁명으로 각광받고 있는 빅데이터, 인공지능 기술을 접목하여 분석을 수행함
- 대다수 선행연구에서 **미대출 도서의 폐기, 이관, 물리적 공간 확대** 등의 접근이 주를 이루고 있으나, 본 연구에서는 **미대출 도서를 인기도서와 연계하여 추천**한다는 점에서 차별성이 있음
- 공간적 범위는 미대출도서가 가장 높게 나타나는 두류도서관으로 한정하였으며, 시간적 범위는 2019년 1월부터 2020년 7월까지의 도서 관련 데이터를 활용함
- 세부적인 데이터로는 두류도서관 도서정보, 대출이력, 회원정보 등을 수집하였으며, 추가적으로 도서관 정보나루의 API를 활용하여 도서 상세정보 등을 수집하였음
- 미대출 도서 추천시스템을 구현하기 위해 크게 2가지 방법으로 시도하였으며, 다음과 같음
- 첫 번째, 도서별 제목과 소개 내용을 결합하여 데이터를 생성하였으며, 이중 명사를 추출하여 Spherical K-means 클러스터링과 **토픽모델링(LDA) 기법**을 사용하여 **카테고리를 추출**함
- 두 번째, **미대출 도서와 인기 도서 사이의 코사인 유사도(Cosine Similarity)**를 계산하여 인기 도서와 유사한 미대출 도서를 성·연령별로 추천하는 콘텐츠 기반 추천시스템을 구현함
- 이를 시각화 프로그램인 Power BI를 통해 웹형태로 구축하여 고도화된 경험을 제공함

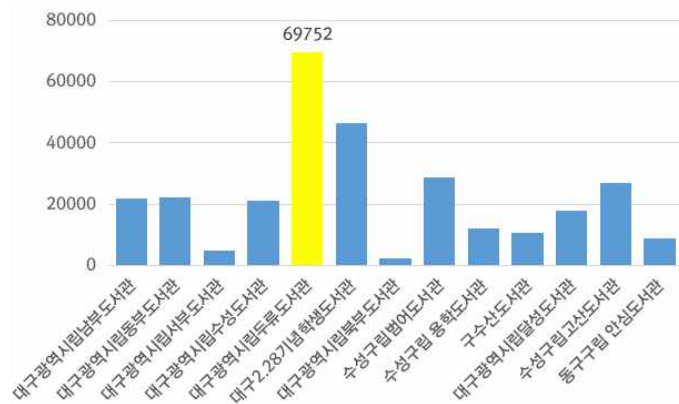
주제어 : 도서관 장서포화, 미대출도서, 추천시스템, 코사인 유사도, 콘텐츠기반

1) 대구디지털산업진흥원 빅데이터활용센터 센터장, 053-655-7050, aut7767@dip.or.kr
 2) 대구디지털산업진흥원 빅데이터활용센터 전임연구원
 3) 경북대학교 통계학과 연구인턴
 4) 경북대학교 수학과 연구인턴
 5) 영남대학교 컴퓨터공학과 연구인턴
 6) 한동대학교 경영학과/데이터사이언스 연구인턴

1 서론

I 연구 필요성 및 목적

- 현재 도서관별 지속적인 신규도서 확보로 공간이 부족하고, 보유도서 대비 편향된 대출 등의 사유로 예산대비 효율성 감소의 문제가 발생함
- 도서관별로 보유중인 도서들 중 대출이 이루어지지 않은 미대출 도서가 전체 도서의 50%를 차지하며 운영 효율성 개선의 필요성이 증대됨
- 대구시 도서관의 운영현황을 살펴보면 <그림 1>과 같이 두류도서관의 경우 미대출도서가 69,752권으로 가장 높게 나타나며, 대구 2·28기념학생도서관, 수성구립 범어도서관 순으로 나타남



<그림 1> 대구광역시 도서관별 미대출 도서 건수

- 본 연구에서는 미대출 도서가 가장 높게 나타나는 두류도서관으로 공간적 범위를 한정하여 장서 포화 문제를 빅데이터 분석으로 접근하고자 하며, 분석 방법론은 추천시스템¹⁾에서 주로 활용되는 콘텐츠 기반 알고리즘을 적용함
- 도서 대출의 경우 성·연령별로 선호하는 카테고리가 다를 것이라 가정하고, 토픽모델링에서 주로 활용되는 군집분석과 LDA 기법을 활용하여 군집별 키워드를 추출하였음
- 추출된 카테고리별 미대출 도서와 대출 도서의 코사인 유사도를 계산하여 인기 도서와 유사한 미대출 도서를 추천해주는 방법으로 미대출 도서 활성화에 기여하고자 함

1) 추천시스템 : 콘텐츠 기반, 협업필터링 알고리즘이 있으며, 최근 이를 혼합한 하이브리드형(기업별 자체 개발)을 적용하는 사례도 있음

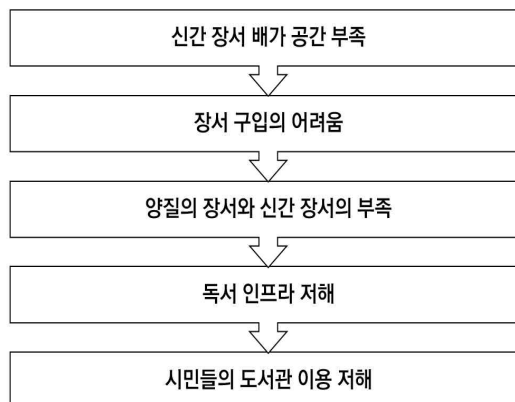
I 공공도서관 장서관리에 관한 선행연구

- 공공도서관 장서관리에 관한 주요 선행연구는 <표 1>과 같으며, 그 외에도 장서관리의 중요성을 강조하는 다양한 선행연구가 존재함

| 연구자(년도) | 주요 내용 |
|-------------|---|
| 윤희윤(2009) | 미래 장서개발의 노선과 전략 제시 |
| 이지연(2009) | 국내 공공도서관의 장서관리 현황을 분석하고 요구사항을 조사 |
| 김선애 등(2009) | 공공도서관의 소장 자료의 분석을 통해 주제별 분포와 노후화 정도를 파악 |
| 이진경 등(2011) | 설문조사를 통한 이용자의 장서 이용 행태 파악 |
| 조윤희(2011) | 공공도서관 장서 폐기 실태 조사를 통한 개선 방안 모색 |
| 김희환(2015) | 장서의 관리와 폐기에 대한 이론적인 부분과 실제적인 측면 연구 |

<표 1> 장서관리에 관한 주요 선행연구

- 공공도서관 장서의 포화에 관한 사례 연구논문(김홍렬, 2019)에서 장서 포화의 원인으로 자료이용 공간 협소, 적정 장서수 산출불가, 장서의 이관이나 제거한 도서를 보존할 제3의 공간 부재, 장서 폐기 절차의 복잡함을 제기함
- 장서 포화의 문제점은 <그림 2>와 같으며, 이를 해결하기 위한 방안으로 이용가치가 낮은 장서의 이관 및 폐기와 도서관 공간의 전면 재조정으로 자료이용 공간의 확대를 제시함



<그림 2> 장서 포화의 문제점

II 연구의 차별성

- 기존의 공공도서관 장서관리에 대한 연구를 살펴보면 대다수가 장서개발, 관리, 폐기에 집중되어 있으며, 그 외에 장서 포화의 원인을 분석하고 정책적 제언을 하는 연구도 있음
- 선행연구에서 장서관리에 대한 정책적 제언 및 해결 방안으로는 이용가치가 낮은 장서의 이관 및 폐기, 공간 확대를 제시하고 있으나, 이는 전통적인 문제 해결 방법으로 판단됨
- 미대출 도서의 이관, 폐기, 공간 확충도 필요하나, 이는 별도의 관리 예산과 운영인력이 소요되는 사항으로 현실적인 한계가 존재함
- 이에 본 연구에서는 기존의 물리적인 접근방법이 아닌 장서의 대출 활성화를 통한 운영 효율화를 진행하고자 하며, 이를 위해 최근 4차 산업혁명으로 각광받고 있는 빅데이터, 인공지능 기술들을 접목하여 분석을 수행함
- 성·연령별로 Spherical K-means 클러스터링과 토픽모델링(LDA)을 통해 카테고리를 분류하고 콘텐츠 기반 알고리즘을 적용하여 인기도서와 유사한 미대출 도서를 추천한다는 점이 기존 연구와 차별성을 지님

3 자료의 수집

I 대구광역시립 두류도서관 데이터

- 미대출 도서 추천 시스템을 구현하고자 공간적 범위를 두류도서관으로 한정하며, 시간적 범위는 2019년 1월부터 2020년 7월까지의 1년 6개월간의 데이터를 활용함
- 제공된 데이터는 크게 3가지로 <표 2>와 같이 분류됨

| 구분 | 분석 자료 |
|-----------------------|---|
| 장서/대출목록 (272,652권) | 번호, 도서명, 저자, 출판사, 발행 연도, ISBN ²⁾ , 세트 ISBN, 부가기호 ³⁾ , 권, 주제분류번호, 도서권수, 대출건수, 등록일자 |
| 대출 이력 (213,093건) | 사용자 키, 도서 키, KDC ⁴⁾ , 도서명, 저자, 출판사, 대출일, 반납일, 미디어 형태 등 |
| 회원 정보 (35,561명) | 사용자 키, 등록일, 관리구분, 관리 도서관, 생년월일, 성별, 우편번호, 도서 대여 횟수, 연체 횟수, 누적 연체 횟수 |

<표 2> 두류도서관 자료의 구성 요소

2) ISBN(International Standard Book Number) : 책을 서점에 유통시키기 위해 발급받는 도서번호(국제표준도서번호)

3) 부가기호 : 도서의 독자대상, 발행형태, 내용 등의 분류를 위해 지정한 기호

4) KDC(Korean Decimal Classification) : 한국십진분류법에 따른 도서 분류키

- 자료의 구성요소를 세부적으로 살펴보면 장서/대출목록의 경우 <그림 3>과 같이 나타나며, 272,652권의 장서에 따른 13개의 자료로 구성됨

| 번호 | 도서명 | 저자 | 출판사 | 발행년도 | ISBN | 세트 ISBN | 부가 기호 | 권 | 주재분류 번호 | 도서 권수 | 대출 건수 | 등록일자 |
|-----|-----------------------------------|--------------------|---------|------|---------------|---------|-------|-----|---------|-------|-------|------------|
| 0 1 | 마이펫의 이중생활=The secret life of pets | 크리스 리노드, 조나단 밀발 감독 | 유니버설픽처스 | 2019 | 8809639452011 | NaN | NaN | 2 | 688.6 | 1 | 0 | 2019-12-26 |
| 1 2 | 엘리펀트킹 덩보 | 하디 모하메디안 감독 | 비디오영향 | 2019 | 8809681640466 | NaN | NaN | NaN | 688.6 | 1 | 0 | 2019-12-26 |
| 2 3 | 덩보 | 팀 버튼 감독 | 홀트디즈니 | 2019 | 8809505724396 | NaN | NaN | NaN | 688.6 | 1 | 0 | 2019-12-26 |
| 3 4 | 캡틴 마블 | 애너 보든, 라이언 플렉 감독 | 홀트디즈니 | 2019 | 8809505723566 | NaN | NaN | NaN | 688.2 | 1 | 0 | 2019-12-26 |
| 4 5 | 메리 포핀스 리턴즈 | 롭 마샬 감독 | 홀트디즈니 | 2019 | 8809505723399 | NaN | NaN | NaN | 688.2 | 1 | 0 | 2019-12-26 |

<그림 3> 장서/대출목록 자료

- 대출이력 자료는 <그림 4>와 같이 나타나며 213,093권의 대출 도서에 따른 14개의 자료로 구성됨

| 이용자 ID | 도서 관리번호 | 도서 청구 기호 | 도서명 | 저자 | 출판사 | 대출일 | 반납 마감일 | 반납일 | 대출 장소 | 대출형태 코드 | 반납형태 코드 | 관리구분 | 매체구분 |
|--------|----------------|----------------|--------------------------------|--------------------------|--------|---------------------|---------------------|---------------------|-------|---------|---------|-------------|------------|
| 0 | 20000071263297 | 90000100169591 | 910 본격 한중일 세계사. 5. 열도의 게임 | 금시니스트 글 그림 | 위즈덤하우스 | 2019-08-14 16:41:37 | 2019-08-28 23:59:59 | 2019-08-28 12:41:18 | NaN | 0 | 0.0 | 대구광역시립두류도서관 | 인쇄자료(책 자형) |
| 1 | 20000071263297 | 90000098290884 | 910 본격 한중일 세계사. 2. 태평천국 라이징 | 금시니스트 글 그림 | 위즈덤하우스 | 2019-08-14 16:41:38 | 2019-08-28 23:59:59 | 2019-08-28 12:41:19 | NaN | 0 | 0.0 | 대구광역시립두류도서관 | 인쇄자료(책 자형) |
| 2 | 20000071263297 | 90000100103983 | 558 (항공우주 과학자가 들려주는) 비행기의 모든 것 | 안석민, 구삼숙, 권기정 지음; 홍원표 그림 | 찰리북 | 2019-07-23 13:23:54 | 2019-08-06 23:59:59 | 2019-08-06 18:16:05 | NaN | 0 | 3.0 | 대구광역시립두류도서관 | 인쇄자료(책 자형) |
| 3 | 20000071263297 | 90000098290884 | 910 본격 한중일 세계사. 2. 태평천국 라이징 | 금시니스트 글 그림 | 위즈덤하우스 | 2019-07-23 13:29:03 | 2019-08-06 23:59:59 | 2019-08-06 18:07:32 | NaN | 0 | 0.0 | 대구광역시립두류도서관 | 인쇄자료(책 자형) |
| 4 | 20000071263297 | 90000099042599 | 910 본격 한중일 세계사. 3. 일본 개황 | 금시니스트 글 그림 | 위즈덤하우스 | 2019-08-14 16:41:37 | 2019-08-28 23:59:59 | 2019-08-28 12:41:17 | NaN | 0 | 0.0 | 대구광역시립두류도서관 | 인쇄자료(책 자형) |

<그림 4> 대출이력 자료

- 회원정보 자료는 <그림 5>와 같으며, 35,561명의 회원에 따른 12개의 자료로 구성됨

| 이용자 ID | 등록일 | 도서관코드 | 도서관 | 생년월일 | 성별 | 우편번호 | 대출도서수 | 연체횟수 | 누적연체횟수 | 생년 | 나이 |
|--------|-----------------|---------------------|----------------|---------------------|----|-------|-------|------|--------|--------|------|
| 0 | 120000003496343 | 2020-01-13 11:30:43 | AB 대구광역시립두류도서관 | 1957-04-25 00:00:00 | F | 42763 | 0.0 | NaN | NaN | 1957.0 | 64.0 |
| 1 | 120000003509102 | 2020-01-13 12:55:25 | AB 대구광역시립두류도서관 | 2004-03-12 00:00:00 | M | 41849 | NaN | NaN | NaN | 2004.0 | 17.0 |
| 2 | 120000003511846 | 2020-01-13 13:26:54 | AB 대구광역시립두류도서관 | 2005-10-28 00:00:00 | F | 41804 | 0.0 | NaN | NaN | 2005.0 | 16.0 |
| 3 | 120000003506086 | 2020-01-13 12:10:30 | AB 대구광역시립두류도서관 | 1981-06-11 00:00:00 | F | 42637 | 0.0 | NaN | NaN | 1981.0 | 40.0 |
| 4 | 120000003280683 | 2020-01-13 10:08:45 | AB 대구광역시립두류도서관 | 1982-04-25 00:00:00 | M | 43016 | 1.0 | NaN | NaN | 1982.0 | 39.0 |

<그림 5> 회원정보 자료

II 도서관 정보나루 API

- 대구시 두류도서관 도서 자료만으로는 고도화된 추천시스템 구현이 어려울 것으로 판단되어, 도서관 정보나루(<https://www.data4library.kr/>)에서 제공하는 Open API를 활용함
- 도서관 정보나루에서는 도서관별 장서/대출, 인기 대출도서, 추천도서, 도서 상세정보, 도서 키워드 목록, 도서별 이용 분석 등의 자료를 API 형태로 제공하며, 본 연구에서는 콘텐츠 기반의 추천서비스 고도

화를 위해 <그림 6>과 같이 도서 상세조회 정보를 호출함

- 도서관 정보나루 API의 경우 일일 제공사용량이 500건으로 제한되어 있어, 이를 별도의 IP등록 작업을 통해 자료호출을 진행함

○ 도서 상세 조회

| | |
|------|---|
| 주요기능 | 빅데이터 분석 플랫폼에서 수집된 서지 및 대출정보를 기반으로 분석한 도서의 상세정보를 제공합니다. ISBN을 통해 특정 도서의 상세정보와 대출정보를 조회하실 수 있습니다. |
| 활용예제 | 아래와 같이 오픈 API 프로토콜 입력 규칙에 따라 베이스URL, 타겟 DB, 질문식, 검색항목등을 조합하여 오픈 API를 호출하면 검색결과를 XML로 제공합니다. (푸른색은 필수입력 항목) ISBN만을 전송할 경우 도서명, KDC, 출판일, 저자, 출판사, 도서이미지 URL, 책소개 정보를 조회하실 수 있으며 추가정보 요청시 해당 도서의 인기대출 지역, 연령, 설명, 대출건수를 제공합니다. |

[http://data4library.kr/api/srchDtlList?authKey=\[발급받은키\]&isbn13=9788983921987&loaninfoYN=Y](http://data4library.kr/api/srchDtlList?authKey=[발급받은키]&isbn13=9788983921987&loaninfoYN=Y)

<그림 6> 도서관 정보나루(도서 상세조회 API)

- 도서의 상세정보를 호출하기 위해 두류도서관에서 보유하고 있는 도서의 ISBN을 기준으로 호출하였으며, 호출된 자료는 <표 3>과 같이 도서명, KDC, 출판일, 저자, 출판사, 도서 이미지 URL, 책 소개 정보 등의 정보를 포함하고 있음

| 구분 | 분석 자료 |
|---------|--|
| 도서 상세조회 | 저자, 도서이미지URL, 책이름, KDC, 책 소개 내용, ISBN(10자리), ISBN(13자리), 출판년, 출판일, 출판사 등 |

<표 3> 도서관 정보나루(도서 상세조회) 자료의 구성 요소

- 출력된 자료의 구조는 <그림 7>과 같이 반정형 데이터인 XML, JSON 형태로 제공되어 데이터 분석이 가능한 데이터프레임 구조로의 변환이 필요함

| | |
|--|--|
| <pre style="font-family: monospace; font-size: 0.9em;"> <no>1</no> - <bookname> - <![CDATA[나미야 잡화점의 기적 -- </pre> | <pre style="font-family: monospace; font-size: 0.9em;"> - <authors> - <![CDATA[지은이: 히가시노 게이고 -- </pre> |
|--|--|

<그림 7> 도서관 정보나루(도서 상세정보) 호출된 데이터(JSON)

- 따라서 파이썬을 활용하여 도서관 정보나루 API로부터 호출된 반정형 데이터를 정형 데이터로 변환 하는 작업을 수행함

| <pre> <![CDATA[책과함께]]> </publisher> <class_no> <![CDATA[910]]> </class_no> <publication_year> <![CDATA[2019]]> </publication_year> <bookImageURL> <![CDATA[https://image.aladin.co.kr/product/19832/23/cover/k062636291_1.jpg]]> </bookImageURL> <isbn> <![CDATA[118899039X]]> </isbn> <isbn13> <![CDATA[9791188990399]]> </isbn13> </pre> | <table border="1"> <thead> <tr> <th>A</th> <th>B</th> <th>C</th> <th>D</th> <th>E</th> </tr> </thead> <tbody> <tr> <td>번호</td> <td>도서명</td> <td>출판일자</td> <td>저자명</td> <td>출판사</td> </tr> <tr> <td>15</td> <td>실크로드</td> <td>2019</td> <td>수전 윗필드</td> <td>책과함께</td> </tr> <tr> <td>16</td> <td>언어의 온도 :말과 글에는 나름의 따뜻함과 자...</td> <td>2016</td> <td>지은이: 이기주</td> <td>말글터</td> </tr> <tr> <td>17</td> <td>설민석의 한국사 대모험 12 - 조선의 발명품...</td> <td>2019</td> <td>설민석, 스토리박스 (지은이), 정현희 (그림...)</td> <td>아이휴먼</td> </tr> <tr> <td>18</td> <td>(설민석의) 한국사 대모험</td> <td></td> <td>글: 설민석,그림: 정현희</td> <td>휴먼큐브</td> </tr> <tr> <td>19</td> <td>(설민석의) 한국사 대모험</td> <td></td> <td>글: 설민석,그림: 정현희</td> <td>휴먼큐브</td> </tr> <tr> <td>20</td> <td>(설민석의) 한국사 대모험</td> <td></td> <td>글: 설민석,그림: 정현희</td> <td>휴먼큐브</td> </tr> <tr> <td>21</td> <td>설민석의 한국사 대모험 11 - 고구려 편...</td> <td>2019</td> <td>설민석, 스토리박스 (지은이), 정현희 (그림...)</td> <td>아이휴먼</td> </tr> <tr> <td>22</td> <td>(설민석의) 한국사 대모험</td> <td></td> <td>글: 설민석,그림: 정현희</td> <td>휴먼큐브</td> </tr> <tr> <td>23</td> <td>(설민석의) 한국사 대모험</td> <td></td> <td>글: 설민석,그림: 정현희</td> <td>휴먼큐브</td> </tr> </tbody> </table> | A | B | C | D | E | 번호 | 도서명 | 출판일자 | 저자명 | 출판사 | 15 | 실크로드 | 2019 | 수전 윗필드 | 책과함께 | 16 | 언어의 온도 :말과 글에는 나름의 따뜻함과 자... | 2016 | 지은이: 이기주 | 말글터 | 17 | 설민석의 한국사 대모험 12 - 조선의 발명품... | 2019 | 설민석, 스토리박스 (지은이), 정현희 (그림...) | 아이휴먼 | 18 | (설민석의) 한국사 대모험 | | 글: 설민석,그림: 정현희 | 휴먼큐브 | 19 | (설민석의) 한국사 대모험 | | 글: 설민석,그림: 정현희 | 휴먼큐브 | 20 | (설민석의) 한국사 대모험 | | 글: 설민석,그림: 정현희 | 휴먼큐브 | 21 | 설민석의 한국사 대모험 11 - 고구려 편... | 2019 | 설민석, 스토리박스 (지은이), 정현희 (그림...) | 아이휴먼 | 22 | (설민석의) 한국사 대모험 | | 글: 설민석,그림: 정현희 | 휴먼큐브 | 23 | (설민석의) 한국사 대모험 | | 글: 설민석,그림: 정현희 | 휴먼큐브 |
|--|---|------|-------------------------------|------|---|---|----|-----|------|-----|-----|----|------|------|--------|------|----|------------------------------|------|----------|-----|----|------------------------------|------|-------------------------------|------|----|----------------|--|----------------|------|----|----------------|--|----------------|------|----|----------------|--|----------------|------|----|----------------------------|------|-------------------------------|------|----|----------------|--|----------------|------|----|----------------|--|----------------|------|
| A | B | C | D | E | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 번호 | 도서명 | 출판일자 | 저자명 | 출판사 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 15 | 실크로드 | 2019 | 수전 윗필드 | 책과함께 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 16 | 언어의 온도 :말과 글에는 나름의 따뜻함과 자... | 2016 | 지은이: 이기주 | 말글터 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 17 | 설민석의 한국사 대모험 12 - 조선의 발명품... | 2019 | 설민석, 스토리박스 (지은이), 정현희 (그림...) | 아이휴먼 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 18 | (설민석의) 한국사 대모험 | | 글: 설민석,그림: 정현희 | 휴먼큐브 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 19 | (설민석의) 한국사 대모험 | | 글: 설민석,그림: 정현희 | 휴먼큐브 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 20 | (설민석의) 한국사 대모험 | | 글: 설민석,그림: 정현희 | 휴먼큐브 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 21 | 설민석의 한국사 대모험 11 - 고구려 편... | 2019 | 설민석, 스토리박스 (지은이), 정현희 (그림...) | 아이휴먼 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 22 | (설민석의) 한국사 대모험 | | 글: 설민석,그림: 정현희 | 휴먼큐브 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 23 | (설민석의) 한국사 대모험 | | 글: 설민석,그림: 정현희 | 휴먼큐브 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

〈그림 8〉 반정형 데이터 예시(좌), 정형 데이터 예시(우)

- 수집된 도서 상세조회 자료는 〈그림 9〉와 같이 나타나며, 기존의 두류도서관 도서정보 자료와 비교하여 도서 이미지, 책 소개 정보 등이 추가로 반영되어 있어 이를 본 연구에 활용하기로 함
- 특히, 도서 소개의 경우 자연어 기반의 100~200자 이내로 작성이 되어 있으며, 이를 자연어 처리를 통해 콘텐츠 기반 추천시스템 고도화에 활용함

| 번호 | 도서명 | 출판일자 | 저자명 | 출판사 | 주제분류 | 출판년도 | 이미지URL | ISBN | 책소개 | |
|----|-----|------------------------------|------|-------------------------------|------|------|--------|--------------------------------|--------------|------------------------------|
| 14 | 15 | 실크로드 | 2019 | 수전 윗필드 의 지음 ;이재철 옮김 | 책과함께 | 910 | 2019.0 | https://image.aladin.co.kr/... | 9.791190e+12 | 세계적 석학 80여 명이 참여하고 7개 언어로... |
| 15 | 16 | 언어의 온도 :말과 글에는 나름의 따뜻함과 자... | 2016 | 지은이: 이기주 | 말글터 | 818 | 2016.0 | http://image.aladin.co.kr/... | 9.791200e+12 | 언어에는 나름의 온도가 있다고 생각하는 저자가... |
| 16 | 17 | 설민석의 한국사 대모험 12 - 조선의 발명품... | 2019 | 설민석, 스토리박스 (지은이), 정현희 (그림...) | 아이휴먼 | NaN | 2019.0 | https://image.aladin.co.kr/... | 9.791190e+12 | 재미있는 스토리와 다양한 한국사 이야기를 통해... |
| 17 | 18 | (설민석의) 한국사 대모험 | NaN | 글: 설민석,그림: 정현희 | 휴먼큐브 | 911 | NaN | http://image.aladin.co.kr/... | 9.791200e+12 | 국민 역사 강사 설민석의 첫 한국사 학습만화... |
| 18 | 19 | (설민석의) 한국사 대모험 | NaN | 글: 설민석,그림: 정현희 | 휴먼큐브 | 911 | NaN | https://bookthumb-phinf.ps... | 9.791200e+12 | 국민 역사 강사 설민석의 첫 한국사 학습만화... |
| 19 | 20 | (설민석의) 한국사 대모험 | NaN | 글: 설민석,그림: 정현희 | 휴먼큐브 | 911 | NaN | https://image.aladin.co.kr/... | 9.791190e+12 | 재미있는 스토리와 다양한 한국사 이야기를 통해... |
| 20 | 21 | 설민석의 한국사 대모험 11 - 고구려 편... | 2019 | 설민석, 스토리박스 (지은이), 정현희 (그림...) | 아이휴먼 | 911 | 2019.0 | https://image.aladin.co.kr/... | 9.791190e+12 | 재미있는 스토리와 다양한 한국사 이야기를 통해... |
| 21 | 22 | (설민석의) 한국사 대모험 | NaN | 글: 설민석,그림: 정현희 | 휴먼큐브 | 911 | NaN | http://image.aladin.co.kr/... | 9.791200e+12 | 국민 역사 강사 설민석의 첫 한국사 학습만화... |
| 22 | 23 | (설민석의) 한국사 대모험 | NaN | 글: 설민석,그림: 정현희 | 휴먼큐브 | 911 | NaN | https://bookthumb-phinf.ps... | 9.791190e+12 | 국민 역사 강사 설민석의 첫 한국사 학습만화... |
| 23 | 24 | (설민석의) 한국사 대모험 | NaN | 글: 설민석,그림: 정현희 | 휴먼큐브 | 911 | NaN | https://image.aladin.co.kr/... | 9.791190e+12 | 재미있는 스토리와 다양한 한국사 이야기를 통해... |

〈그림 9〉 도서관 정보나루 호출된 데이터(CSV)

4 데이터 전처리

I 도서 상세 정보 데이터 전처리

- 두류도서관 도서 상세정보의 경우 데이터의 잡음이 일부 존재하여 이를 제거하여 분석을 수행함
- R을 활용하여 도서 상세 정보 데이터에서 도서 소개가 공백인 행을 제거하고 도서명이 동일한 도서의 경우 도서 소개를 통합하여 전처리를 수행함

| 번호 | 도서명 | 출판일자 | 저자명 | 출판사 | 주제분류 | 출판년도 | 이미지URL | ISBN | 책소개 |
|----|---------------------|------|---------------------|------|------|--------|---------------------|--------------|--------------------|
| 12 | Why? 교과서 중심 사회 | NaN | 글: 남준자, 최익규 | 예림당 | 300 | NaN | NaN | 9.788930e+12 | NaN |
| 13 | Why? 교과서 중심 수학 | NaN | 글: 김문수, 최병국 | 예림당 | 410 | NaN | NaN | 9.788930e+12 | NaN |
| 14 | Why? 교과서 중심 국어 | NaN | 글: 윤상석, 한종전 | 예림당 | 710 | NaN | NaN | 9.788930e+12 | NaN |
| 15 | 실크로드 | 2019 | 수전 핏필드 외 지음, 이재황 옮김 | 책과함께 | 910 | 2019.0 | https://image.al... | 9.791190e+12 | 세계적 석학 80여 명이 장... |
| 16 | 언어의 온도 :말과 글에는 나... | 2016 | 지은이: 이기주 | 말글터 | 818 | 2016.0 | http://image.ala... | 9.791200e+12 | 언어에는 나름의 온도가 있다... |
| 17 | 설민석의 한국사 대모험 12 ... | 2019 | 설민석, 스토리박스 (지은이...) | 아이휴먼 | NaN | 2019.0 | https://image.al... | 9.791190e+12 | 재미있는 스토리와 다양한 한... |
| 18 | (설민석의) 한국사 대모험 | NaN | 글: 설민석,그림: 정현희 | 휴먼큐브 | 911 | NaN | http://image.ala... | 9.791200e+12 | 국민 역사 강사 설민석의 첫... |
| 19 | (설민석의) 한국사 대모험 | NaN | 글: 설민석,그림: 정현희 | 휴먼큐브 | 911 | NaN | https://bookthum... | 9.791200e+12 | 국민 역사 강사 설민석의 첫... |
| 20 | (설민석의) 한국사 대모험 | NaN | 글: 설민석,그림: 정현희 | 휴먼큐브 | 911 | NaN | https://image.al... | 9.791190e+12 | 재미있는 스토리와 다양한 한... |

〈그림 10〉 도서 상세정보 데이터 전처리(전)

| 번호 | 도서명 | 출판일자 | 저자명 | 출판사 | 주제분류 | 출판년도 | 이미지URL | ISBN | 책소개 |
|----|---------------------|------|---------------------|------|------|--------|---------------------|--------------|--------------------|
| 15 | 실크로드 | 2019 | 수전 핏필드 외 지음, 이재황 옮김 | 책과함께 | 910 | 2019.0 | https://image.al... | 9.791190e+12 | 세계적 석학 80여 명이 장... |
| 16 | 언어의 온도 :말과 글에는 나... | 2016 | 지은이: 이기주 | 말글터 | 818 | 2016.0 | http://image.ala... | 9.791200e+12 | 언어에는 나름의 온도가 있다... |
| 17 | 설민석의 한국사 대모험 12 ... | 2019 | 설민석, 스토리박스 (지은이...) | 아이휴먼 | NaN | 2019.0 | https://image.al... | 9.791190e+12 | 재미있는 스토리와 다양한 한... |
| 18 | (설민석의) 한국사 대모험 | NaN | 글: 설민석,그림: 정현희 | 휴먼큐브 | 911 | NaN | http://image.ala... | 9.791200e+12 | 국민 역사 강사 설민석의 첫... |

〈그림 11〉 도서 상세정보 데이터 전처리(후)

- 토픽모델링인 LDA를 수행하기 위해 데이터의 숫자, 특수문자, 기호 등을 제거하는 텍스트 클렌징 작업을 수행하였으며, 전처리 전/후는 〈표 4〉와 같이 나타남

| 전처리 전 | | 전처리 후 | |
|--------|--|--------|---|
| 0 | 언어의 온도 :말과 글에는 나름의 따뜻함과 차가움이 있다 언어에는 나름의 온도가 ... | 0 | 언어의 온도 말과 글에는 나름의 따뜻함과 차가움이 있다 언어에는 나름의 온도가 있... |
| 1 | [빅북] 피자를 먹지 마! 토도의 그림책 시리즈. 피자들이 자신을 먹지 말라고 부탁하... | 1 | 빅북 피자를 먹지 마 토도의 그림책 시리즈 피자들이 자신을 먹지 말라고 부탁하는 이... |
| 2 | [빅북] 사탕괴물 - 생활습관 (매, 춤치, 간식) 주인공의 입안 가득 춤치 군... | 2 | 빅북 사탕괴물 생활습관 매 춤치 간식 주인공의 입안 가득 춤치 군이 살고 있는... |
| 3 | [빅북] 웃음이 퐁퐁퐁 바람 그림책 76권. 언제나 웃음 띤 얼굴로 즐거웠던 아기... | 3 | 빅북 웃음이 퐁퐁퐁 바람 그림책 권 언제나 웃음 띤 얼굴로 즐거웠던 아기 되지 퐁... |
| 4 | [빅북] 한숨 구멍 새 유치원 등원을 앞둔 아이의 불안하고 걱정스러운 마음을 세밀하... | 4 | 빅북 한숨 구멍 새 유치원 등원을 앞둔 아이의 불안하고 걱정스러운 마음을 세밀하게 ... |
| 159127 | 안한필 건강교실·천하를 얻고도 건강을 잃으면 무슨 소용이 있겠는가 제독(除毒)·자연... | 159127 | 안한필 건강교실천하를 얻고도 건강을 잃으면 무슨 소용이 있겠는가 제독除毒·자연식·문... |
| 159128 | 되는 집안은 가지나무에 수박 열린다 슬기로운 선인들의 지혜와 교훈을 담은 현양대 국... | 159128 | 되는 집안은 가지나무에 수박 열린다 슬기로운 선인들의 지혜와 교훈을 담은 현양대 국... |
| 159129 | 가스실 아무리 파렴치한 살인마라도 다시는 돌아올 수 없는 가스실로 보내는 것만이 법... | 159129 | 가스실 아무리 파렴치한 살인마라도 다시는 돌아올 수 없는 가스실로 보내는 것만이 법... |
| 159130 | 마왕퇴의 귀부인 격변하는 문화혁명기에 세계를 뒤집은 고고학 발굴의 혁명, 마왕퇴 발... | 159130 | 마왕퇴의 귀부인 격변하는 문화혁명기에 세계를 뒤집은 고고학 발굴의 혁명 마왕퇴 발굴... |
| 159131 | 눈물로 남은 너 2 김량 감성소설. 지치고 고단한 영혼에 새로운 생명을 불어넣기도 ... | 159131 | 눈물로 남은 너 김량 감성소설 지치고 고단한 영혼에 새로운 생명을 불어넣기도 하고... |

〈표 4〉 텍스트 클렌징 작업(전)/(후)

II 대출기록, 회원정보 데이터 전처리

- 성·연령별 인기 카테고리 도서를 추천하기 위해 대출기록과 회원정보 데이터 결합이 필요하며, 분석에 필요한 회원번호, 등록일, 생일, 성별, 출생연도, 나이를 제외한 데이터는 삭제함
- 이상치와 결측치를 제거하기 위해 출생연도가 2020년 이후인 데이터와 누락 데이터들을 일괄적으로 삭제함
- 이산형의 출생연도 데이터를 범주형(10대 이하, 20대, 30대, 40대, 50대, 60대 이상)으로 전처리

하여 연령대 변수를 생성함

- 전처리가 완료된 데이터는 기존 35,561개에서 33,551개로 감소하였으며, <그림 12, 13>과 같이 나타나며, 대출 정보 데이터 셋을 구축함

```
> summary(user)
  USER_KEY      REG_DATE      BIRTHDAY      GENDER      year      age
Length:35561   Length:35561   Length:35561   Length:35561   Min.   :1900   Min.   : -26.0
Class :character Class :character Class :character Class :character 1st Qu.:1970   1st Qu.: 29.0
Mode  :character Mode  :character Mode  :character Mode  :character Median:1981   Median: 40.0
Mean  :1980   Mean  : 41.4
3rd Qu.:1992   3rd Qu.: 51.0
Max.  :2047   Max.  :121.0
NA's  :5      NA's  :5
```

<그림 12> 대출기록, 회원정보 데이터 전처리(전)

```
> summary(user)
  USER_KEY      REG_DATE      BIRTHDAY      GENDER      year      age
Length:33551   Length:33551   Length:33551   Length:33551   Min.   :1900   Min.   :  1.00
Class :character Class :character Class :character Class :character 1st Qu.:1970   1st Qu.: 29.00
Mode  :character Mode  :character Mode  :character Mode  :character Median:1981   Median: 40.00
Mean  :1980   Mean  : 40.82
3rd Qu.:1992   3rd Qu.: 51.00
Max.  :2020   Max.  :121.00
```

<그림 13> 대출기록, 회원정보 데이터 전처리(후)

- 대출기록, 회원정보를 USER_KEY를 기준으로 결합하여 <그림 14>와 같이 대출 정보 데이터를 구축함

| ISBN | USER_KEY | CLASS_NO | TITLE | AUTHOR | PUBLISHER | LOAN_DATE | RETURN_PLAN_DATE | RETURN_DATE | LOAN_TYPE_CODE | RETURN_TYPE_CODE | MANAGE_CODE | MEDIA_CODE | REG_DATE | BIRTHDAY | GENDER | year | age | age_group |
|------|------------|--------------|--|------------|-----------|---------------------|------------------|---------------------|----------------|------------------|-----------------|---------------|---------------------|----------------------|--------|------|-----|-----------|
| 0 | NaN | 9.000000e+13 | 그림 만 하는 누가 다 읽 었 을 까 | 박원서 지 | 웅진출판 | 2019-09-22 14.22 | 2019-10-06 23.59 | 2019-10-13 16.04 | 0 | 0.0 | 대구광역시립독 후도서관 | 인쇄자료(학자 형) | 2009-02-19 12.17 | 1973-09- 13.00.00 | F | 1973 | 48 | 4 |
| 1 | NaN | 9.000000e+13 | 讀 山. 7 | 朴夢九 [著] | 自由文庫社 | 2019-06-08 15.09 | 2019-06-22 23.59 | 2019-06-21 16.03 | 0 | 0.0 | 대구광역시립독 후도서관 | 인쇄자료(학자 형) | 2017-02-19 16.07 | 1966-05- 02.00.00 | F | 1966 | 55 | 5 |
| 2 | 8985062638 | 9.000000e+13 | 한 말 살 았 던 하 소 서 | 박원서 지 | 윙출판사 | 2019-07-11 12.55 | 2019-08-09 23.59 | 2019-08-07 12.04 | 0 | 0.0 | 대구광역시립독 후도서관 | 인쇄자료(학자 형) | 2010-12-16 12.24 | 1943-05- 07.00.00 | M | 1943 | 78 | 6 |
| 3 | 8985062638 | 9.000000e+13 | 한 말 살 았 던 하 소 서 | 박원서 지 | 윙출판사 | 2019-03-22 14.38 | 2019-04-05 23.59 | 2019-04-05 12.13 | 3 | 3.0 | 대구광역시립독 후도서관 | 인쇄자료(학자 형) | 2018-10-12 13.18 | 1975-01- 19.00.00 | F | 1975 | 48 | 4 |
| 4 | 8985062638 | 9.000000e+13 | 한 말 살 았 던 하 소 서 | 박원서 지 | 윙출판사 | 2019-11-23 15.28 | 2019-12-07 23.59 | 2019-12-07 17.57 | 0 | 0.0 | 대구광역시립독 후도서관 | 인쇄자료(학자 형) | 2016-04-06 18.40 | 1968-05- 10.00.00 | F | 1968 | 53 | 5 |

<그림 14> 대출 정보 데이터

III 도서 상세 데이터와 대출 정보 데이터 병합

- 도서 대출, 미대출 도서를 분류하기 위해서는 도서 상세 데이터와 대출 정보 데이터 병합이 필요하며, 이를 ISBN, 도서명 등을 기준으로 병합함
- 대출 정보 데이터의 경우 ISBN이 없거나 10자리, 또는 이상치인 도서가 포함되어 있어 이를 다음과 같은 방식으로 처리함
- ISBN이 있는 도서의 경우, 10자리인 ISBN은 13자리로 변환하고 추가 이상치 제거를 진행함
- ISBN이 없는 도서의 경우, 두 데이터의 도서 제목의 공백과 특수문자를 제거한 후 두 데이터 간의 제목의 유사도를 비교하여 유사도가 높은 도서의 제목은 도서 상세 데이터에 있는 제목으로 통일하여

파일을 병합함

| 제목 비교 전 | | | | 제목 비교 후 | | | |
|-----------|-----|----|------|--------------|----------|-----|---------------|
| 도서명 | 매칭율 | 번호 | ISBN | 도서명 | 매칭율 | 번호 | ISBN |
| 昆蟲分類學 | | 1 | | 형 빈 중앙 | 1 | 6 | 9788946413399 |
| 변경1不斑의세필1 | | 2 | | 변경 | 1 | 14 | 9788932015125 |
| 변경1不斑의세필1 | | 3 | | 변경 | 1 | 15 | 9788932015125 |
| 변경2不斑의세필2 | | 4 | | 로마인이야기 | 0.923077 | 64 | 9788935654833 |
| 변경3不斑의세필3 | | 5 | | 로마인이야기 | 0.923077 | 65 | 9788935654833 |
| 형빈중앙 | | 6 | | 블랙 커넥션 3 - P | 0.923077 | 79 | 9788912071929 |
| 계룡산16권 | | 7 | | 블랙 커넥션 3 - P | 0.923077 | 80 | 9788912071929 |
| 계룡산16권 | | 8 | | 블랙 커넥션 3 - P | 0.923077 | 81 | 9788912071929 |
| 계룡산16권 | | 9 | | 블랙 커넥션 3 - P | 0.923077 | 82 | 9788912071929 |
| 계룡산16권 | | 10 | | 패스트 | 1 | 191 | 9791186877616 |
| 계룡산16권 | | 11 | | 패스트 | 1 | 192 | 9791186877616 |
| 계룡산16권 | | 12 | | 패스트 | 1 | 193 | 9791186877616 |
| 계룡산16권 | | 13 | | 패스트 | 1 | 194 | 9791186877616 |
| 계룡산16권 | | | | 패스트 | 1 | 195 | 9791186877616 |

〈표 5〉 ISBN이 없는 도서에 대한 제목의 유사도 비교

- 도서상세 데이터와 대출 정보 데이터를 병합하여 대출 도서 목록과 미대출 도서 목록 파일을 생성함

IV 문장 토큰화

- 텍스트 분석을 실시하기 위해서는 컴퓨터가 학습할 수 있도록 형태소⁵⁾로 분리하는 토큰화 (Tokenization) 작업이 별도로 필요함
- 토큰화는 문장 토큰화와 단어를 토큰으로 분리하는 단어 토큰화로 구분되는데, 본 연구에서는 단어 토큰화를 사용하여 분석을 실시함
- 한글의 경우 형태소를 분리할 때 띄어쓰기와 조사로 인해 의미 있는 단어 추출에 어려움이 따르며, 이러한 문제를 해결하기 위해 파이썬 기반의 KoNLPy 라이브러리를 활용함
- 〈표 6〉과 같이 형태소 분리 후 그 중 명사의 품사를 가지고 있는 형태소만 추출함

| 구분 | 내용 |
|----|---|
| 목록 | 언어, 온도, 말, 글, 나름, 움, 언어, 나름, 온도, 생각, 저자, 일상, 발견, 의미, 말, 글, 권, 책, 저자, 단어, 어원, 유래, 언어, 움, 글감, 것, 예찬, 언어, 나름, 온도, 생각, 저자, 일상, 발견, 의미, 말, 글, 권, 책, 저자, 단어, 어원, 유래, 언어, 움, 글감, 것, 예찬, 빅북, 피자, 마, 토토, 그림책, 시리즈, 피자, 자신, 말, 부탁, 이야기, 그림책, 캐릭터, 감각, 색깔, 재미, 아이, 상상력, 호기심, 자극, 이보, 피자, 그림책, 빅북, 사탕, 괴물, 생활, 습관, 때, 충치, 간식, 주인공, 입안, 가득, 충치, 균, 살, 일방, 교훈, 이야기, 달리, 아이, 입장, 접근, 아이, 마음, 구성은, 사탕, 괴물, 의, 가장, 장점, 이자, 독창, 점, 빅북, 웃음, 풍풍풍, 바람, 그림책, 권, 웃음, 먼, 얼굴, 아기, 돼지, 풍풍이, 오늘, 잔뜩, 화가, 모자, 바람, 때문, 잔뜩, 골, 난, 얼굴, 두더지, 동동, 깔깔, 바다, 제안, 빅북, 한숨, 구멍, 새, 유치원 등 |

〈표 6〉 추출된 명사 목록(예시)

5) 형태소 : 단어로서의 의미를 가지는 최소 단위

V 불용어 제거

- 텍스트 분석에서는 데이터 내 유의미한 단어 토큰만을 선별하기 위해 크게 의미가 없는 단어 토큰 즉, 불용어를 제거하는 작업이 필요함
- 단어의 의미가 없다는 것은 언어의 특성상 자주 발생하고, 의미가 낮은 조사 위주의 단어들을 말하는 것으로 KoNLPy에서 기본적으로 제공하는 불용어 사전만을 적용하는 것은 한계가 있음
- 따라서 본 연구에서는 고도화된 텍스트 전처리를 수행하기 위해 <표 7>과 같이 자체적으로 불용어 사전을 생성하였으며, 해당 사전에 포함되는 단어는 총 286개로 구성됨

| 구분 | 내용 |
|----|---|
| 목록 | 우리, 마지막, 실었다, 얼마나, 파악, 스스로, 핵심, 쓰기, 구조, 동시, 최근, 형태, 운영, 적용, 기자, 일반, 공개, 사람, 여기, 진짜, 내용, 주제, 특성, 형성, 공개, 포함, 실용, 그것, 사례, 행동, 문학상, 비밀, 더욱, 존재, 단어, 도전, 기법, 기록, 노력, 출판, 바탕, 위기, 등단, 고민, 혁신, 하루, 메시지, 스토리, 프로젝트, 비교, 전체, 사이, 개인, 년대, 도움, 지은이, 세기, 단어, 시선, 산업, 각각, 기록, 자리, 제작, 지금, 이름, 배경, 진실, 개념, 활동, 정도, 선정, 현실, 인식, 서로, 년도, 나라, 사실, 정보, 시한, 개정판, 민국, 표현, 전달, 효과, 사고, 생활, 자기, 현상, 거리, 가지, 이용, 박사, 시간, 집필, 자료, 작가, 직접, 흐름, 완성, 최고, 자신, 깊이, 소재, 주요, 기초, 별로, 다섯, 순간, 모두, 습관, 처음, 구성, 동안, 문장, 설명, 시각, 모습, 각종, 대한, 오늘, 분야, 대상, 고대, 이번, 하나, 보고 등 |

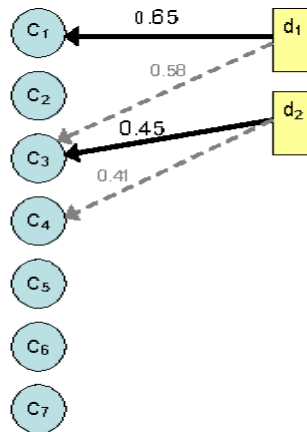
<표 7> 불용어 사전 목록(예시)

5 Spherical K-means 클러스터링 및 토픽모델링 (LDA)

I Spherical K-means 클러스터링

- K-means 알고리즘은 분류예측 분석기법으로 주어진 데이터를 k개의 클러스터로 묶는 알고리즘이며 각 클러스터와 거리 차이의 분산을 최소화하는 방식으로 동작함
- 그 중에서도, Spherical K-means 알고리즘은 코사인 거리를 계산하는 K-means 알고리즘으로 Sparse vector로 표현되는 문서 집합에 사용하기 적절함
- K-means 알고리즘은 다음과 같은 순서로 클러스터링을 수행함
 - 1) 초기 k 평균값은 데이터 오브젝트 중에서 무작위로 뽑힘
 - 2) 각 데이터 오브젝트들은 가장 가까이 있는 평균값을 기준으로 묶임

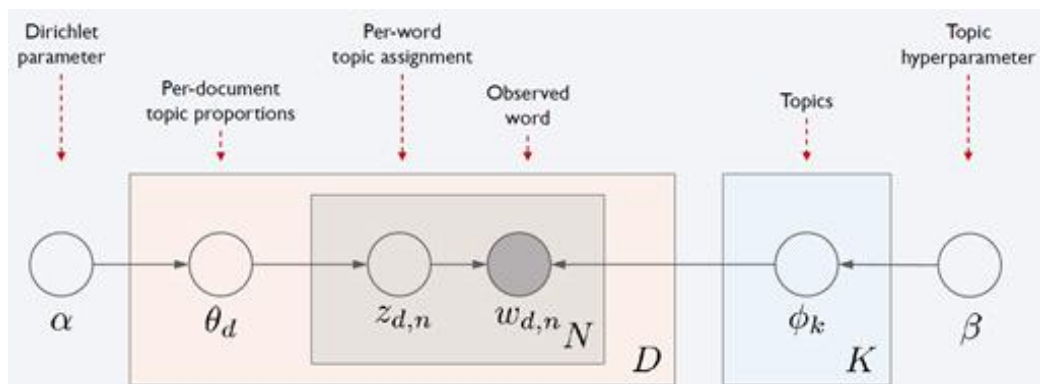
- 3) k개의 클러스터의 중심점을 기준으로 평균값이 재조정됨
- 4) 수렴할 때까지 2, 3번 과정을 반복함
- Spherical K-means 알고리즘은 이러한 클러스터링이 수렴될 때까지 반복적으로 수행됨
- 본 연구에서는 Spherical K-means 클러스터링을 통해 159,132개의 도서 텍스트 데이터를 총 20개의 군집으로 분류하고 각 도서마다 해당하는 군집 이름을 부여함
- Spherical K-means 알고리즘은 <그림 15>와 같은 프로세스로 진행됨



<그림 15> Spherical K-means

II 토픽 모델링 (LDA)

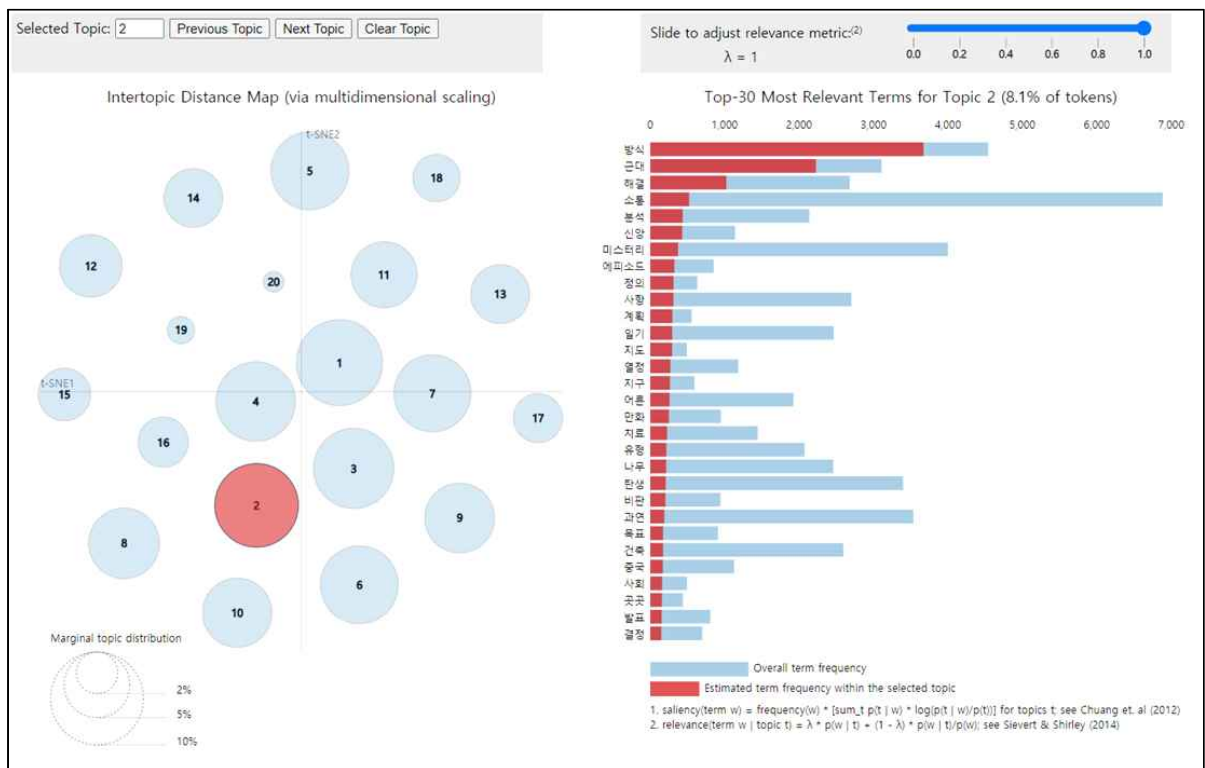
- 잠재 디리클레 할당(Latent Dirichlet allocation, LDA)은 주어진 문서에 대하여 각 문서에 어떤 주제들이 존재하는지를 서술하는 확률적 토픽 모델 기법 중 하나임
- 미리 알고 있는 주제별 단어 수 분포를 바탕으로, 주어진 문서에서 발견된 단어 수 분포를 분석함으로써 해당 문서가 어떤 주제들을 함께 다루고 있을지를 예측할 수 있음



<그림 16> 토픽 모델링(LDA) 알고리즘

- Spherical K-means 클러스터링으로 분류한 데이터를 활용하여 토픽 모델링(LDA) 시각화를 수행하였음

- 그 결과, <그림 17>과 같이 인터랙티브 시각화⁶⁾로 구현하여, 20개의 주제에 각각 포함되는 상위 30개의 단어를 확인할 수 있음
- 세부적으로 살펴보면 시각화 좌측에 위치한 각 원은 도서의 주제를 의미하며, 이는 문학/여행, 종교/철학, 패션/디자인 등이 있으며, 원 안의 숫자는 토픽번호를 나타냄
- 특정한 원을 클릭하게 되면 원은 빨간색으로 표시되고, 우측에는 선택된 주제와 가장 관련 있는 단어 30개를 Bar graph 형태로 출력함
- Bar graph 내 하늘색 그래프는 전체 주제에서의 단어 빈도를 의미하며, 빨간색 그래프는 선택한 주제에서의 단어 빈도를 의미함



<그림 17> 토픽 모델링(LDA) 시각화 결과

III 분석 결과

- 앞서 수행한 군집분석과 토픽모델링을 기반으로 두류도서관내 도서에 대한 20개의 토픽 및 관련 키워드를 <표 8>과 같이 추출함
- 해당 토픽라벨을 대출 도서 목록과 미대출 도서 목록 파일에 추가하여 군집별 의미를 부여함

6) 인터랙티브(interactive) 시각화 : 사용자의 동작에 반응하는 시각화

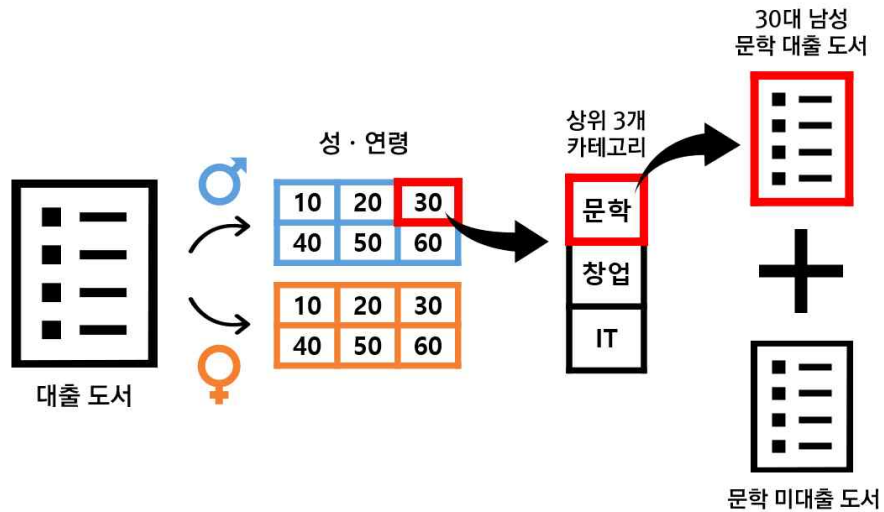
| 순번 | 토픽 | 주요 키워드 |
|----|------------|---|
| 1 | 문학, 여행 | 한국어, 조선시대, 문학, 역사 문화, 기행, 대하 |
| 2 | 사회문제 | 근대, 해결, 소통, 분석, 정의, 사항, 비판, 사회 |
| 3 | 심리, 정치 | 인지, 국가, 서양, 치유, 만남, 가정, 영혼, 상황 |
| 4 | 종교, 철학 | 소통, 미스터리, 원리, 원칙, 세계, 신앙, 종교, 성경 |
| 5 | 건강, 요리 | 인간, 만들기, 나무, 마을, 기적 |
| 6 | 외교, 세계정치 | 고찰, 상황, 프랑스, 평화, 일본, 자유, 역사 문화, 세계사, 세상 |
| 7 | 청소년 교육 | 음악, 한국어, 수업, 관찰 |
| 8 | 취미, 생활, IT | 사진, 비법, 만들기, 도시, 일기, 현대인, 원리, 설계 |
| 9 | 외국어 | 창작, 만남, 프로그램, 친구, 집중, 크게, 국가, 교양, 능력 |
| 10 | 패션/디자인 | 그녀, 인물, 시도 |
| 11 | 의학, 세계소설 | 의학, 세계, 대안, 창작, 연애, 미스터리, 연구, 발표 |
| 12 | 자기계발, 창업 | 탄생, 방식, 비법, 원칙, 일기, 열정, 문제, 분석 |
| 13 | 소설 | 미스터리, 인기, 국민, 스포츠 |
| 14 | 세계사, 역사 | 인문, 국가, 서양, 유럽, 해결, 이론, 리더 |
| 15 | 문화/예술 | 단편, 스타일, 영어, 러시아, 북한, 고전, 종교, 시대, 미술 |
| 16 | 교육 도서 | 준비, 아기, 자녀, 인터넷, 탄생, 어른, 비법, 아동, 방식 |
| 17 | 산문, 시집 | 풍경, 나무, 재미, 일기 |
| 18 | 과학 | 미스터리, 과연, 실험, 유형, 분석, 탄생, 발달 |
| 19 | 만화, 오락 | 전략, 학습, 그리스, 미스터리, 게임, 인터넷 |
| 20 | 한국 수필 | 통찰, 이야기, 상황, 언어, 친구 |

〈표 8〉 토픽별 키워드(20개 토픽)

6 콘텐츠 기반 추천시스템

I 성·연령별 인기 카테고리 추출

- 전처리된 대출 도서 목록 파일에서 대출이력을 기반으로 성·연령별로 분리하여 총 12개의 파일을 생성함
- 각 파일에서 군집 라벨을 기준으로 대출 건수가 많은 상위 3개 카테고리를 추출하였으며, 대출 도서와 미대출 도서 사이의 코사인 유사도를 계산하기 위해 각 카테고리별 대출도서와 미대출 도서를 병합함



〈그림 18〉 성·연령별 인기 카테고리 추출 프로세스

II 코사인 유사도

- 텍스트 마이닝에서 도서 간의 유사도를 계산할 때 일반적으로 코사인 유사도(Cosine Similarity)를 사용하며, 코사인 유사도는 내적공간의 두 벡터 간 각도의 코사인 값을 이용하여 측정된 벡터간의 유사한 정도를 의미함
- 코사인 유사도는 벡터의 크기가 아닌 방향의 유사도를 판단하는 목적으로 사용되며, 두 벡터의 방향이 완전히 같을 경우 1, 90°의 각을 이룰 경우 0, 180°로 완전히 반대 방향인 경우 -1의 값을 가짐
- 속성 A, B의 벡터 값이 각각 주어졌을 때, 코사인 유사도 $\cos(\theta)$ 는 벡터의 스칼라 곱과 크기로 〈그림 19〉와 같이 표현할 수 있음



〈그림 19〉 코사인 유사도 값(예시)

- 코사인 유사도의 계산식은 〈그림 20〉과 같이 나타나며, 코사인 유사도 외에도 도서 간 유사도 비교

방법에는 유클리디안(Euclidean Distance) 유사도, 자카드 유사도(Jaccard Similarity) 등이 있음

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

〈그림 20〉 코사인 유사도 계산식

- 본 연구에서는 추천 시스템에서 가장 보편적으로 사용되는 코사인 유사도를 사용하였으며, 유클리디안 유사도는 두 문서가 이루는 각도를 무시하기에 정확도가 떨어지고, 자카드 유사도의 경우 연산속도 저하, 오류에 대한 알고리즘 자체 필터링 기능이 없기에 휴먼 에러에 취약하다는 단점이 있음

III 피처 벡터화 및 토큰화

- 인기 도서와 미대출 도서 간 코사인 유사도를 계산하기 위해 자연어를 숫자형 피처로 변환하는 피처 벡터화를 수행함
- 피처 벡터화란 비정형 텍스트 데이터를 단어 기반의 다수의 피처로 추출하고, 이를 단어 빈도수와 같은 숫자 값을 부여한 후, 텍스트는 단어의 조합인 벡터 값으로 표현하는 것을 말함
- 피처 벡터화를 수행하는 방식에는 2가지가 있음
 - 1) 카운트 기반 벡터화
 - 2) TF-IDF(Term Frequency - Inverse Document Frequency)
- 첫 번째, 카운트 기반 벡터화는 문맥 안에 단어가 동시에 등장하는 횟수를 세는 방법으로, 〈그림 21〉과 같이 동시 등장 횟수를 하나의 행렬로 나타낸 뒤 그 행렬을 수치화해서 단어 벡터로 만드는 방식임
- 그러나 문서 내용이 방대할 경우 카운트만 계산한다면 매우 높은 빈도로 나타나거나 문장에서 일반적으로 자주 사용되는 단어에도 높은 값을 부여하게 되어, 이러한 문제를 해결하기 위해 TF-IDF(Term Frequency - Inverse Document Frequency)가 개발됨

| | the | red | dog | cat | eats | food |
|--------------------|-----|-----|-----|-----|------|------|
| 1. the red dog → | 1 | 1 | 1 | 0 | 0 | 0 |
| 2. cat eats dog → | 0 | 0 | 1 | 1 | 1 | 0 |
| 3. dog eats food → | 0 | 0 | 1 | 0 | 1 | 1 |
| 4. red cat eats → | 0 | 1 | 0 | 1 | 1 | 0 |

〈그림 21〉 카운트 기반 벡터화(예시)

- 두 번째, TF-IDF는 카운트 기반 벡터화의 단점을 개선하기 위해 다른 문서에는 적게 등장하나 해당 문서에서는 자주 등장하는 단어에 높은 점수를 부여하는 방식으로 동작함
- TF-IDF에서 TF는 Term Frequency의 약자로 문서에서 해당 단어가 나타는 빈도를 의미하며, IDF는 Inverse Document Frequency의 약자로 DF(Document Frequency)의 역수임
- DF는 전체 문서들에서 몇 개의 문서에 해당 단어가 나타나있는지에 대한 값이며, TF-IDF는 이 TF값과 IDF값을 곱한 값이며, 이에 대한 수식은 〈그림 22〉와 같음

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$$

$tf_{i,j}$ = number of occurrences of i in j
 df_i = number of documents containing i
 N = total number of documents

〈그림 22〉 TF-IDF 벡터화 수식

IV 분석 결과

- 전처리된 도서별 피쳐 매트릭스를 활용하여 동일 성·연령 내의 대출 도서와 미대출 도서간 코사인 유사도를 〈그림 23〉과 같이 산정하였음
- 미대출 도서와 대출 도서 사이의 유사도의 평균이 높은 순서대로 추천 순위를 산정하였음

카테고리별 대출 도서

| | 엘리펀트킹 | 덤보 | 캡틴마블 | ... | 실크로드 | 언어의 온도 | mean |
|-------------------|----------|----------|------|----------|----------|----------|------|
| 세상에서 가장 특별한 새 이야기 | 0.008416 | 0.006328 | ... | 0.000000 | 0.002561 | 0.003750 | |
| 과학 용어 사전(증보 제3판) | 0.002835 | 0.000000 | ... | 0.000000 | 0.000000 | 0.001344 | |
| ... | ... | ... | ... | ... | ... | ... | |
| 아마도 세상은 | 0.000000 | 0.018997 | ... | 0.000000 | 0.008418 | 0.001009 | |
| 어린이를 위한 배 세계사 100 | 0.004246 | 0.003998 | ... | 0.000000 | 0.000000 | 0.002017 | |

<그림 23> 도서간 평균 코사인 유사도

- 앞서 산정된 도서간 코사인 유사도를 기반으로 각 성·연령별 코사인 유사도가 높은 도서 100권을 내림차순으로 <그림 24>와 같이 정렬하여 데이터 셋을 구축함

| ISBN | 도서명 | 저자명 | 책소개 | 이미지URL | 주제분류 | 군집 번호 | 분류 | text | 성별 | 연령대 | 카테고리순가중치 | 순위 | 군집 이름 | |
|---------------|--------------|---------------|-------------|------------------------|------------|-------|--------|---------|--------|-----|----------|----------|----------|----------|
| 8809154157149 | 비밀의 화자 | 박렛 올리 | 1909년에 | http://image.aladin.cc | | 6 | 미대출 | 비밀의 화자 | F | 10 | 1 | 100 | 1 청소년교육 | |
| 9788996314271 | 6학년은 왕글쓰기 | 노 주니어 | 노 주니어 | http://ima | 813.7 | 6 | 미대출 | 6학년은 왕글 | F | 10 | 1 | 99 | 2 청소년교육 | |
| 9788986032215 | 적성따라 | 이덕환 | 그 아이들을 | http://image.aladin.cc | | 6 | 미대출 | 적성따라 | F | 10 | 1 | 98 | 3 청소년교육 | |
| 9788951506932 | 지각대장 | 동화창의 | 아이들을 | http://ima | 813.6 | 6 | 미대출 | 지각대장 | F | 10 | 1 | 97 | 4 청소년교육 | |
| 9788951507540 | (꼬마돼지) 동화창의 | 아이들을 | http://ima | 813.6 | 6 | 미대출 | (꼬마돼지) | F | 10 | 1 | 96 | 5 청소년교육 | | |
| 9788985768696 | (보람이랑 김보람) | 김서울 경희 | https://bo | 810.82 | 6 | 미대출 | (보람이랑 | F | 10 | 1 | 95 | 6 청소년교육 | | |
| 9788951506949 | 꽃밭에 내 | 동화창의 | 아이들을 | http://ima | 813.6 | 6 | 미대출 | 꽃밭에 내 | F | 10 | 1 | 94 | 7 청소년교육 | |
| 9788969150592 | (똥똥지 선 지은이) | 김 그림책 학 | https://ima | 375.1 | 6 | 미대출 | (똥똥지 선 | F | 10 | 1 | 93 | 8 청소년교육 | | |
| 9788951508929 | 생일나무 | 동화창의 | 아이들을 | http://ima | 813.6 | 6 | 미대출 | 생일나무 | F | 10 | 1 | 92 | 9 청소년교육 | |
| 9788994545530 | (진짜 진짜 리처드 드 | 아이들의 | https://bo | 556 | 6 | 미대출 | (진짜 진짜 | F | 10 | 1 | 91 | 10 청소년교육 | | |
| 9788932318288 | 나는 아이 | 매건 다 | <뉴욕 E | http://ima | 332.24 | 6 | 미대출 | 나는 아이 | F | 10 | 1 | 90 | 11 청소년교육 | |
| 9788909056915 | 은따나라 | 엄기원 (지 여섯 | 분의 | http://image.aladin.cc | | 6 | 미대출 | 은따나라 | F | 10 | 1 | 89 | 12 청소년교육 | |
| 9788963055428 | (슈퍼맘 박 지은이) | 박 <말똥세 | 중 | https://bo | 720 | 6 | 미대출 | (슈퍼맘 박 | F | 10 | 1 | 88 | 13 청소년교육 | |
| 9788993870206 | 기동찬, 공 서지 | 원 글 | 글담어린 | 0 | http://ima | 710 | 6 | 미대출 | 기동찬, 공 | F | 10 | 1 | 87 | 14 청소년교육 |
| 9788951507557 | 내가 어른 | (동화창의 | 아이들을 | http://ima | 813.6 | 6 | 미대출 | 내가 어른 | F | 10 | 1 | 86 | 15 청소년교육 | |
| 9791187427858 | 내가 엄마 | (한승민 (지 | 바람숲그 | http://image.aladin.c | | 6 | 미대출 | 내가 엄마 | F | 10 | 1 | 85 | 16 청소년교육 | |
| 9788932917085 | 생쥐와 친 | (지은이: 루 | <연애 소 | http://ima | 873 | 6 | 미대출 | 생쥐와 친 | F | 10 | 1 | 84 | 17 청소년교육 | |
| 9788971541036 | 청소년 살 | (이진우 지 | <나 열 | https://bo | 370.4 | 6 | 미대출 | 청소년 살 | F | 10 | 1 | 83 | 18 청소년교육 | |
| 9788983890856 | 다섯가지 | (편집부 (여 | 여느 작은 | http://image.aladin.cc | | 6 | 미대출 | 다섯가지 | F | 10 | 1 | 82 | 19 청소년교육 | |
| 9788950965358 | 생각이 크 | (글: 김윤경, < | 생각이 | https://bo | 1 | 6 | 미대출 | 생각이 크 | F | 10 | 1 | 81 | 20 청소년교육 | |

<그림 24> 미대출 추천도서 목록(예시 : 10대 이하/여성)

- 이를 파이썬 기반의 추천시스템으로 구현하였으나, 일반 사용자가 쉽게 이용할 수 있도록 시각화 프로그램인 Power BI를 활용하여 웹 형태로 구축하였음
- 미대출 도서 추천시스템 UI는 <그림 25>와 같이 구성하였으며, 특정 성·연령을 선택하면 해당 범주의 카테고리 내 미대출 도서 중 코사인 유사도 기반으로 최대 100권의 도서를 추천함
- 추천 받은 도서의 세부적인 정보 검색을 원할 경우 도서 이미지를 클릭하면 화면 하단에 도서에 대한 세부정보가 나타남
- 슬라이서를 통해 도서 출판년도 기간을 선택하여 검색할 수 있고, 이는 미대출 도서 중 최근에 출판된

도서의 비율이 높기 때문에 미대출 도서 대출을 활성화하기 위해 마련한 기능임

| 군집 이름 | 저자명 | 도서명 | 연령대 | 책소개 |
|-------|----------|----------------|-----|---|
| 문학/여행 | 김영훈 지음 | (김영훈 소설집)파라노이아 | 20대 | 전문가의 소설집, 표제작 파라노이아를 비롯하여 어둠의 긴 통로 그 너머면, 87년 여름에는, 추억 속으로, 한계령을 넘어가면, 강릉에서, 까치이야기, 흙 |
| 문학/여행 | 신경숙 지음 | (신경숙 소설집)말기발 | 20대 | 85년 문예중앙으로 등단한 90년대 최고의 여류작가의 소설집. 존재의 고뇌와 육체적 감각의 발달함을 모색한 단편들로 삼의 슬픔, 사랑과 죽음과 청는 정소, 어떤 여자 등 6편의 단편 수록. |
| 문학/여행 | 엄창섭 지음 | (엄창섭 소설)황금색 발톱 | 20대 | 1990년 동아일보 신춘문예로 등단한 작가의 소설집. 성과 권력에 대한 작가 나름의 발문 색칠하는 여자를 비롯해 저능아 민태와 행철이, 나가 주축이 색 발톱소설기계 등 8편을 엮었다. |
| 문학/여행 | 김시습 외 지음 | (필독)우리고전소설 | 20대 | 조선 조의 한문소설인 사씨남정기부터 장화홍련전, 심청전, 배비장전, 한중록, 은세계, 금수회의록 등 신소설에 이르기까지 각 작품별 작가 소개, 줄거리, 성격을 정리하고 그 분석하는 |

〈그림 25〉 미대출 도서 추천시스템 구현(시각화)

7 결론

1 연구의 정책적 시사점

- 본 연구에서는 도서관 장서 포화문제를 빅데이터, 인공지능 분야에서 각광받고 있는 추천시스템 알고리즘을 활용하여 문제해결에 접근하였음
- 이를 위해 콘텐츠 기반 알고리즘을 활용하여 미대출 도서 추천시스템을 구현하였으며, 도출된 시사점은 다음과 같음
- 첫 번째, 도서관의 주요 이슈인 편향된 대출 이력, 운영 효율화 등의 문제를 미대출 도서 추천시스템을 통해 일부 개선할 수 있을 것으로 판단됨
- 두 번째, 구현된 시스템을 키오스크, 이동도서관, 스마트 도서관에서의 IT기술과 결합하여 사용자들에게 고도화된 경험을 제공할 수 있을 것으로 기대함
- 세 번째, 베스트셀러, 출판사의 마케팅 등의 인기도서 위주의 대출이력에서 그동안 조명 받지 못

했던 도서를 발굴함으로써 도서 산업에 일부 기여를 할 수 있을 것으로 사료됨

- 네 번째, 추후 진행되는 대구광역시 공공도서관 추천시스템 구축 시 미대출 도서도 함께 추천함으로써 고도화된 추천시스템 구현이 가능할 것으로 판단됨
- 마지막으로 본 연구는 대구광역시립 두류도서관을 기반으로 구현하였고, 이를 토대로 대구광역시 내 다른 공공 도서관에도 적용한다면 도서관 운영 활성화에 기여할 것으로 기대되며, 해당 시스템을 활용하여 추가 인사이트 도출도 가능할 것으로 예상됨

II 연구의 한계

- 본 연구의 분석결과를 바탕으로 연구의 한계는 다음과 같음
- 첫 번째, 도서 소개 내용의 경우 출판사들의 마케팅 정보가 다수 포함되어 있어, 실제 도서 내용과는 일부 차이가 발생할 수 있음
- 두 번째, 대구시 공공도서관 데이터의 경우 이상치, 결측치가 상당수 존재하여 해당 부분의 처리 방법과 범위에 따라 추천시스템의 성능이 상이할 것으로 판단됨
- 세 번째, 성·연령별 군집분석을 통해 키워드를 추출하고, 군집별 인기 카테고리 내 미대출도서와 대출도서의 유사도 평균을 활용하여 도서를 추천하는 방법론에 있어 다소 논리적인 근거가 부족함
- 마지막으로 본 연구는 장서 포화도에 따른 운영 효율화에 대해서는 기여할 수 있을 것으로 판단되나, 도서관 내 부족한 장서의 면적 등에 관한 물리적인 문제 해결에 대해서는 한계점을 가지는 것으로 판단됨

I 대시보드 UI (Power BI)



대시보드 ① : 콘텐츠 기반 미대출 도서 추천 화면

- 성·연령별 도서 카테고리 선택 가능
- 도서의 출판년도 선택 가능



대시보드 ① : 성·연령을 선택한 화면

- 특정 성·연령별 카테고리 설정 시 우측에 도서의 이미지 표시, 하단에는 미대출 도서 리스트 표시
- 카테고리 선택 하단에 도서 출판년도 설정 시 출판년도에 해당하는 미대출 도서 이미지, 리스트 표시



대구시립두류도서관 회원정보

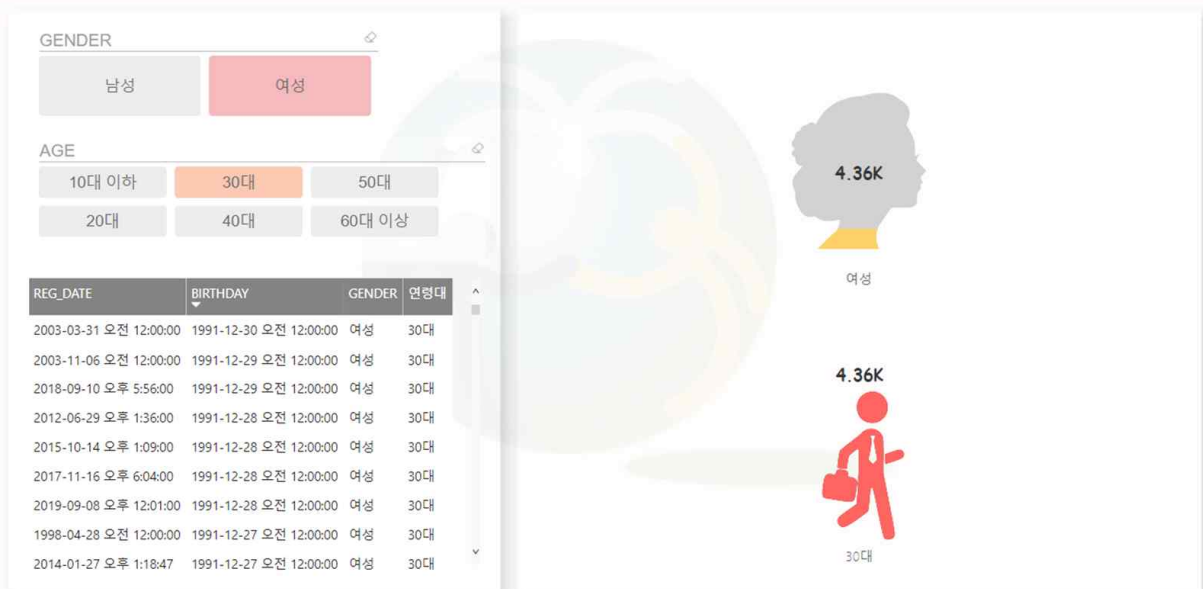


대시보드 ② : 도서관 회원 정보

- 두류도서관의 회원 분포를 인포그래픽으로 시각화
- 좌측에서 성·연령 선택 가능



대구시립두류도서관 회원정보



대시보드 ② : 특정 성·연령을 선택한 화면

- 선택한 성·연령에 따라 우측의 인포그래픽 시각화

대구시립두류도서관 도서정보

책소개

파이썬 머신러닝·최신 예측 분석 기반으로 깊은 통찰력을 얻는 머신러닝 필수 가이드
9791187497035

파이썬이 데이터 과학과 관련된 하나의 이유를 설명하고 있다.

(저를 만나서) 머신러닝과 딥러닝 :C언어로 구현하며 익히는 기본 개념
9788965401681

머신러닝의 다양한 분야를 알기 쉽게 설명하고 이를 토대로 딥러닝을 살펴본다. 단순한 개념에 맞는 간단한 프로그램을 직접 구현해봄으로써 이 기술이 어떻게 사용되는지 이해할 수 있다.

(파이썬을 이용한) 머신러닝, 딥러닝 실전 개발 입문 : 웹 크롤링과 스크레이핑부터 머신러닝-딥러닝까지 체계적으로 배우기
9791158390679

머신러닝의 바탕이 되는 데이터를 수집하고, 수집된 데이터를 기반으로 머신러닝을 수행하는 과정을 소개한다. 가용된 데이터를 활용하여 딥러닝을 구현하는 방법과 딥러닝을 적용하는 방법까지 소개한다.

파이썬 라이브러리를 활용한 머신러닝 :사이킷런 핵심 개발자가 쓴 머신러닝과 데이터 과학 실무서
9788966483394

머신러닝과 인공지능에 대한 사전 지식이 필요 없는 입문서로, 복잡한 수학을 통쾌하지 않고 실용적으로 머신러닝을 공부하는 모든 단계를 설명한다. 파이썬과 사이킷런을 통해 머신러닝을 구현하는 방법과 딥러닝을 적용하는 방법까지 소개한다.

| ISBN | 도서명 | 책소개 |
|---------------|--|---|
| 9791188059812 | (가장 쉬운) AI(인공지능) 입문서 :머신러닝, 딥러닝에 대한 원리를 알기 쉽게 설명 | 인공지능의 원리에 대해 쉽게 이해할 수 있도록 다양한 사례를 들어 설명하고, 미래에 있다. 특히 인공지능의 핵심 이론인 머신러닝, 딥러닝에 대한 이해를 높이기 위해 다양한... |
| 9791158390761 | (상생한 현장의 실무 예제로 배우는) 리얼월드 머신러닝 | 실무 개발자에게 머신러닝 프로젝트 실행 기술을 가르치고자 고안된 실용 가이드. 확률 머신러닝 시스템을 성공적으로 구축 및 배치하려고 준비하는 독자를 위해 생생한 현장... |
| 9791158390419 | (알고리즘으로 배우는) 인공지능, 머신러닝, 딥러닝 입문 | 위키백스 데이터 사이언스 시리즈 2권. 이 책은 인공지능에 관심이 있는 모든 독자를 위해, 특히 머신러닝과 딥러닝 분야의 개발자들에게는 쉽게 설명된 핵심 이론과 함께 적용... |
| 9788965401681 | (저를 만나서) 머신러닝과 딥러닝 :C언어로 구현하며 익히는 기본 개념 | 머신러닝의 다양한 분야를 알기 쉽게 설명하고 이를 토대로 딥러닝을 살펴본다. 단순한 이에 맞는 간단한 프로그램을 직접 구현해봄으로써 이 기술이 어떤 것인지를 쉽게 이... |
| 9791158390679 | (파이썬을 이용한) 머신러닝, 딥러닝 실전 개발 입문 | 머신러닝의 바탕이 되는 데이터를 수집하고, 수집된 데이터를 기반으로 머신러닝을 수행하는 과정을 소개한다. 가용된 데이터를 활용하여 딥러닝을 구현하는 방법과 딥러닝을 적용하는 방법까지 소개한다. |

ISBN

도서명

저자명

대시보드 ③ : 도서관 전체 도서 정보, 목록 화면

- 도서관에 존재하는 전체 도서 정보와 책 소개를 워드클라우드로 시각화
- ISBN, 도서명, 저자명을 이용해 도서 검색 가능

대구시립두류도서관 도서정보

책소개

(저를 만나서) 머신러닝과 딥러닝 :C언어로 구현하며 익히는 기본 개념
88965401681

머신러닝의 다양한 분야를 알기 쉽게 설명하고 이를 토대로 딥러닝을 살펴본다. 단순한 개념에 맞는 간단한 프로그램을 직접 구현해봄으로써 이 기술이 어떻게 사용되는지 이해할 수 있다.

(파이썬을 이용한) 머신러닝, 딥러닝 실전 개발 입문 : 웹 크롤링과 스크레이핑부터 머신러닝-딥러닝까지 체계적으로 배우기
9791158390679

머신러닝의 바탕이 되는 데이터를 수집하고, 수집된 데이터를 기반으로 머신러닝을 수행하는 과정을 소개한다. 가용된 데이터를 활용하여 딥러닝을 구현하는 방법과 딥러닝을 적용하는 방법까지 소개한다.

파이썬 라이브러리를 활용한 머신러닝 :사이킷런 핵심 개발자가 쓴 머신러닝과 데이터 과학 실무서
9788966483394

머신러닝과 인공지능에 대한 사전 지식이 필요 없는 입문서로, 복잡한 수학을 통쾌하지 않고 실용적으로 머신러닝을 공부하는 모든 단계를 설명한다. 파이썬과 사이킷런을 통해 머신러닝을 구현하는 방법과 딥러닝을 적용하는 방법까지 소개한다.

| ISBN | 도서명 | 책소개 |
|---------------|-------------------------------|---|
| 9791158390679 | (파이썬을 이용한) 머신러닝, 딥러닝 실전 개발 입문 | 머신러닝의 바탕이 되는 데이터를 수집하고, 수집된 데이터를 기반으로 머신러닝을 수행하는 과정을 소개한다. 가용된 데이터를 활용하여 딥러닝을 구현하는 방법과 딥러닝을 적용하는 방법까지 소개한다. |

ISBN

도서명

저자명

대시보드 ③ : 특정 책을 선택한 화면

- 특정 책을 선택하면 ISBN과 도서명 그리고 책 소개를 좌측 하단에 표시
- 책 소개를 워드클라우드로 시각화

대구시립두류도서관 도서정보

책소개

자바를 처음 만나는 자바스크립트
JavaScript
185890777

자바를 만나는 자바스크립트와 JavaScript의 기본을 탄탄히 다지고 상세한 예제와 실용적인 코드로 구성한 고재, 기본 소개는 물론 프로그래밍에 필요한 사고방식에도 중점을 두고 설명하고 있으며, 하는 기능들이 어디서 어떻게 사용되는지도 쉽게 알 수 있도록 다양한 예제를 소개한다.

자바 for 안드로이드 프로그래밍
9788956747439

(조보자도 쉽게 배우는) 자바 for 안드로이드 프로그래밍 : 기초 자바를 통해 안드로이드 앱을 쉽게 개발에 도전한다

안드로이드 앱 개발에 필요한 자바 문법과 안드로이드의 구조적 특징을 이해하여 안드로이드 앱을 쉽게 개발하는 방법을 설명한다. 특히, 다양한 예제를 통해 쉬운 문법도 완성도 높고, 어려운 문법도 쉽게 배울 수 있도록 다양한 예제를 소개한다.

자바 세상의 빌드를 이끄는 메이븐
9788979148138

(자바 세상의 빌드를 이끄는) 메이븐

실제 프로젝트로 학습하는 생생한 메이븐 활용 전략이 담긴 책. 상당수의 자바 오픈 소스 프로젝트가 메이븐 빌드를 주로 사용하지만, 이를 현업에 적용하는 것을 어려워하는 개발자를 위해 기획되었다. 메이븐을 처음 접하는 개발자도 두루 활용할 수 있도록 다양한 예제를 소개한다.

리눅스 바로 배우는 자바
9788966488627

(당장 현장에 투입된 조보 개발자를 위한) 실무에서 바로 통하는 자바 Java 8 대응

이 책은 실무에 곧바로 뛰어들어야 하는 조보 개발자를 위한 책입니다. 현장 개발자로 활동하며 신입사원을 교육한 저자가 실무에 꼭 필요한 내용을 담았습니다. 자바 핵심 문법, 프로그래밍 기법, 라이브러리, API 등 실무에 꼭 필요한 내용을 소개합니다.

자바를 배우는 자바스크립트
9788979148261

저자의 강의 중 집중했던 자바적으로 많이 소개된 글라서 소개할 예제를 소개하고 실용성과 동기를 설명한다. 프로그래밍 기법, 라이브러리, API 등 실무에 꼭 필요한 내용을 소개합니다.

| ISBN | 도서명 | 책소개 |
|---------------|--|---|
| 9788970507699 | (HTML과 자바 스크립트 중심으로) 홈페이지 만들기 =Homepage design | HTML과 자바 스크립트 중심으로 『홈페이지 만들기』. 이 책은 짧은 시간에 자신을 알리 웹 문서 기반인 HTML 개요를 설명하였고 DYNAMIC HTML, 스타일, 필터 효과 등을 서술 객체, COOKIE를... |
| 9788979142068 | (IT Expert) 모바일 자바 프로그래밍 | 자바 개발자들을 대상으로 J2ME-J2EE 연동을 통한 실전 모바일 프로그래밍 방법을 제시 (http://www.mobilejava.co.kr) 대표 권기경씨와 한국 자바 개발자 협의회(www.javacom) |
| 9788979142037 | (IT Expert) 자바 개발자를 위한 XML 프로그래밍 | 자바언어의 기초를 익힌 프로그래머들을 대상으로 자바나 JSP를 이용해 XML 애플리케이션을 이용한 XML 프로그래밍 방법을 살펴봄. 꼭 필요한 내용을 간결하고 정확하게 설명한 |
| 9788966187270 | (가장 빨리 만나는) 자바 8 | 2014년 4월 출시된 자바 8(Java SE 8, JDK 8)의 주요 기능을 주제별로 정리해서 알려준 API, 함수형 프로그래밍의 개념을 비유있게 다룬다. |
| 9788966260928 | (개발자가 반드시 알아야 할) 자바 성능 최적화 | 이 책은 고성능 애플리케이션을 위해 고려해야 할 복잡한 요소와 성능 개선 방법을 쉽고 |

대시보드 ③ : 도서명으로 검색 시 화면

- 입력한 도서명을 포함하는 도서 목록 표시, 위드클라우드 시각화

대구시립두류도서관 인기도서 순위

대출건수

35 | 79

82년생 김지영 : 조남주 장편소설
9788937473135

오늘의 젊은 작가 13권 조남주 장편소설, 시사 교양 프로그램에서 10년 동안 일한 방송작가답게 서민들의 일상에서 발생하는 비극을 사실적이고 감동대 울음 스토리로 표현하는 그의 문체는 감동과 울림이 넘친다.

언어의 온도
9791195522125

언어는 나와의 온도가 있다고 생각하는 저자가 일상에서 발견한 의미 있고 맑고 아름다운 글의 정에 매달다. 저자는 단어의 어원과 유래, 그런 언어가 지닌 자기애와 따뜻함을 글감 삼아 허탈해보이는 것들의 신중함을 애써한다. 언어

나는 나로 살기로 했다
9791187119845

우리가 온전한 '나'로 살아가기 위해 무엇이 필요한지 말해주는 책이다. 돈 많고 잘나가는 타인의 SNS를 뜯어보며 비참해질 필요 없고, 스스로에게 변명하고 모두에게 이해받으려 애쓰지 말고 이야기한다. 무엇이든 무조건 열심히 할

둘이 될 수 없는 약속 : 야쿠마루 가쿠 장편소설
9788998274795

제51회 에드가와랑표상 수상 작가 야쿠마루 가쿠 장편소설 '너리 추리소설, 주인공이 과거에 저지른 죄, 그리고 15년 전에 했던 어떤 약속을 둘러싸고 벌어지는 미스터리 소설이다. (1권) 이후 2, 3권도 기대하고 있다

무례한 사람에게 웃으며 대처하는 법 : 인생 자체는 긍정적으로 개소리에는 단호하게
9788936434267

감자기 선을 뚫고 감정의 통로로 오셨어요??? 기분은 어떻게? 기분은 어떻게? 기분은 어떻게?

| ISBN | 대출건수 | TITLE | AUTHOR | 주제 |
|---------------|------|---|-----------------------------|----------|
| 9788937473135 | 79 | 82년생 김지영 : 조남주 장편소설 | 조남주 지음 | 사회문제 |
| 9791195522125 | 69 | 언어의 온도 | 이기주 지음 | 문화 예술 |
| 9791187119845 | 64 | 나는 나로 살기로 했다 | 글:그림: 김수현 | 만화, 오락 |
| 9788998274795 | 53 | 둘이 될 수 없는 약속 : 야쿠마루 가쿠 장편소설 | 야쿠마루 가쿠 지음; 김성미 옮김 | 문화, 여행 |
| 9788957369364 | 48 | 무례한 사람에게 웃으며 대처하는 법 : 인생 자체는 긍정적으로 개소리에는 단호하게 | 정문정 지음 | 자기계발, 창업 |
| 9788936434267 | 46 | 아몬드 : 손원평 장편소설 | 지은이: 손원평 | 소설 |
| 9788901223032 | 46 | 하마터면 열심히 살 뻔했다 : 야매 독도 에세이 | 하완 지음 | 자기계발, 창업 |
| 9791162332542 | 45 | (주리 천재) 영명이 탄생. 5. 유적에서 날아온 SOS 사건 | 트름글:그림: 김정화 옮김 | 종교, 철학 |
| 9788934972464 | 45 | 사피엔스 : 유인원에서 사이보그까지. 인간 역사의 대담하고 위대한 질문 | 유발 하라리 [지음]; 조현욱 옮김; 이태수 감수 | 세계사, 역사 |
| 9788901219943 | 45 | 신경 끄기의 기술 | 마크 맨슨 지음; 한재호 옮김 | 자기계발, 창업 |
| 9788972756194 | 43 | 나미야 잡화점의 기적 | 히가시노 게이고 지음; 양윤옥 옮김 | 문화, 여행 |

대시보드 ④ : 인기 도서 순위 화면

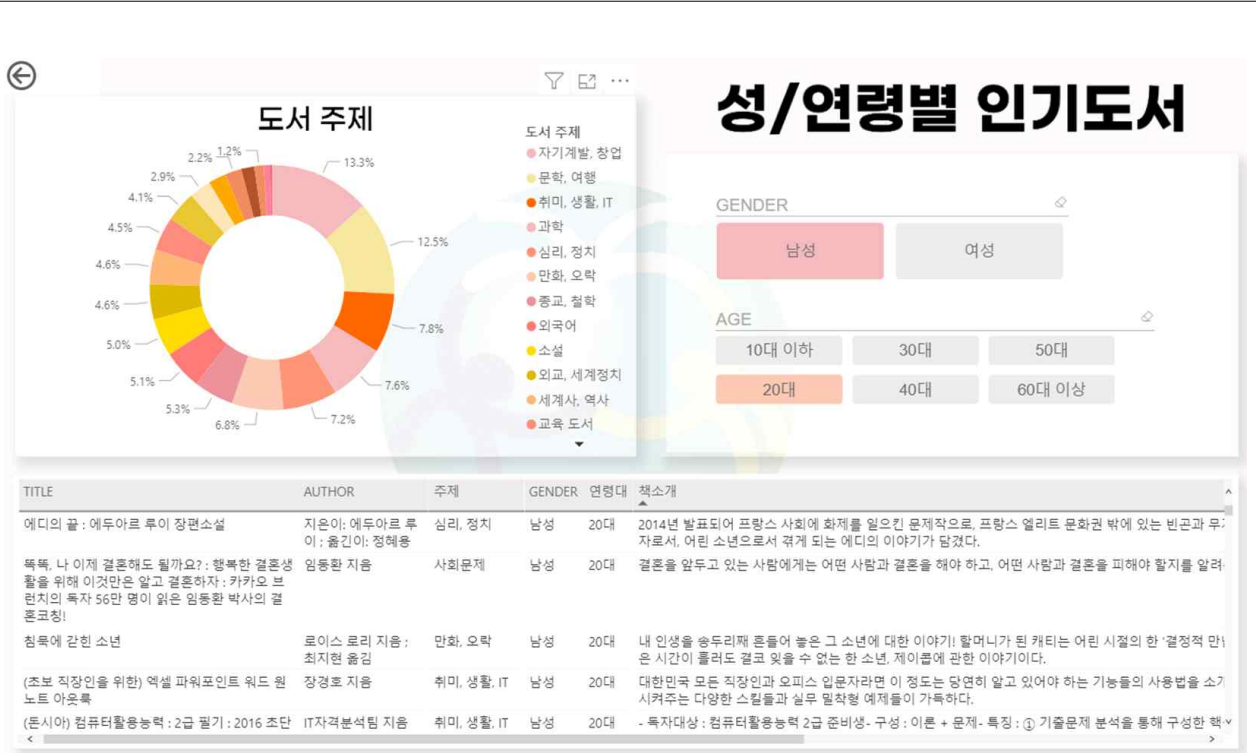
- 대출 건수 범위를 선택하면 그 범위에 해당하는 도서관 인기 도서 목록 표시

- 24 -



대시보드 ⑤ : 성·연령별 인기 도서

- 우측 상단에서 성·연령을 선택할 수 있음
- 좌측에는 선택한 성·연령에 따른 도서의 주제 분포를 도넛형 그래프로 시각화
- 하단에 성·연령별 인기 도서 목록 표시



대시보드 ⑤ : 성·연령 선택 시 화면

- 선택한 성·연령에 따라 도서 주제 분포 그래프 시각화 및 도서 목록 표시

참고 문헌

1. 김건욱 외 1인, 대구디지털산업진흥원, 빅데이터보고서, 2020.09
콘텐츠, 협업필터링을 활용한 공공도서관 도서 추천시스템 구현
 2. 박대우, 송실대학교 박사학위 논문,
빅데이터 기반 도서추천시스템 구축을 위한 아키텍처에 관한 연구
 3. 김태영외 2인, 한국도서관·정보학회지 제49권 제2호,
빅데이터 로그 기반 도서관 이용자 및 대출 현황 분석
 4. 이은지 외 2인, 디지털도서관 제80호
도서관 빅데이터 기반 서비스 개발 및 운영 방안 연구
 5. 백지연 외 3인, 한국도서관·정보학회지 제49권 제4호
빅데이터 기반 도서관 어린이청소년서비스 현황분석 및 개선방안
 6. 최혜봉, 한국정보기술학회 제14권 제9호
지역도서관을 위한 도서 추천 인공 신경망 모델
 7. 박철민, 파이썬 머신러닝 완벽가이드
 8. 도서관정보정책위원회, 제3차 도서관발전종합계획
-

본 보고서는 대구의 경제·사회 전망을 빅데이터로 진단하고 대응방안 제시, 다양한 분석기법들을 소개하는 분석자료입니다.

본 자료의 내용은 연구자의 견해이며 대구디지털산업진흥원의 공식 견해와 반드시 일치하는 것은 아닙니다.