

DCMI 2021: Best Practice

Implementation and challenge to improve metadata workflow in National Library of Korea

Yoonkyung CHOI
National Library of Korea

2021. 10. 5.

Content

- Metadata workflow of NLK

- Project 1 |

Automatic subject indexing using Korean bookstore's data

- Project 2 |

Metadata automatic conversion of journal articles based on external data

Metadata workflow

- Cataloging is a mix of simple and intellectual tasks



- Traditional cataloging process is still maintained
 - Most of the work is done manually by ourselves and outsourced
- Usually, description is simple and easy
 - It takes time because it has to be filled out one by one
- In particular, subject analysis is the most expensive
 - Accuracy varies widely from person to person

글자크기 16 px 단축키 ▲(Ctrl+Y) ↔(Ctrl+U) 한자를 한글로 변환

001	KM0201654352 ▲
005	20170313162731 ▲
007	ta ▲
008	161012s2016 ulk 000 kor ▲
020	▼a9788959064120▼g03300:▼c#15000 ▲
023	▼aCIP2016022207 ▲
052	01▼a331.1▼b16-30 ▲
056	▼a331.1▼26 ▲
082	01▼a302▼223 ▲
245	00▼a생각과 착각 /▼d강준만 지음 ▲
260	▼a서울 :▼b인물과사상사,▼c2016 ▲
300	▼a387 p. ;▼c23 cm ▲
490	10▼a세상을 꿰뚫는 50가지 이론 ;▼v5 ▲
650	8▼a사회 심리[社會心理] ▲
700	1▼a강준만,▼g康俊晩,▼d1956-▼OKAC201211150 ▼
830	0▼a세상을 꿰뚫는 50가지 이론 ;▼v5 ▲
950	0▼b#15000 ▲
049	0▼LEM6513097
	▼LEM6513098▼c2 ▲↔

```
<?xml version="1.0" encoding="UTF-8"?>
<mods xsi:schemaLocation="http://www.loc.gov/mods/v3 http://www.loc.gov/standards/mods/v3/mods-3-5.xsd"
xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:ndklids="http://nl.go.kr/schema/license/terms/" xmlns="http://www.loc.gov/mods/v3">
  <titleInfo>
    <title>빅데이터 시대의 개인정보보호 자율규제 활성화 방안</title>
  </titleInfo>
  <titleInfo type="parallel">
    <title>study on self-regulation improvement of personal information protection in the era of big data</title>
    <nonSort>A</nonSort>
  </titleInfo>
  <name type="personal" usage="primary">
    <namePart>채항석</namePart>
    <displayName>Chae, Hyang-Seok</displayName>
    <alternativeName altType="no specific type">
      <namePart>Chae, Hyang-seok</namePart>
    </alternativeName>
  </name>
  <typeOfResource>텍스트</typeOfResource>
  <genre>학술논문</genre>
  <originInfo>
    <place>
      <placeTerm type="text">서울</placeTerm>
      <placeTerm type="code" authority="kormarccountry">ulk</placeTerm>
    </place>
    <publisher>법학연구원</publisher>
    <dateIssued encoding="iso8601">201706--</dateIssued>
    <issuance>연속간행자료</issuance>
    <frequency>간행빈도 불명</frequency>
  </originInfo>
  <language>
    <languageTerm type="code" authority="iso639-2b">kor</languageTerm>
  </language>
  <physicalDescription>
    <form>전자자료(Application)</form>
    <reformattingQuality>access</reformattingQuality>
    <internetMediaType>application/pdf</internetMediaType>
    <extent>40 p.</extent>
    <digitalOrigin>born digital</digitalOrigin>
  </physicalDescription>
  <abstract type="초록">최근 빅데이터를 활용한 서비스의 확대로 개인정보의 수집 및 이용이 증대하여 개인정보보호 조치의 요구도 높
    은 경직된 정부규제만으로 전문적이고 신속하게 대응할 수 없는 문제가 있다. 이에 개인정보보호 분야는 민간부문에 의한 자율규제
```

Offline resources

- monograph, serial, journal article, multimedia, etc.
- described based on Korean Cataloging Rule, 4th edition(KCR4) and Korean MARC(KORMARC)
- approximately 150,000 data per year

Online resources

- e-book, e-journal, e-article, multimedia, etc.
- described based on KCR4 and MODS(Metadata Object Description Scheme)
- approximately 200,000 data per year

Project 1

⦿ Purpose

- Developing a pilot service that automatically recommends topical terms
- The term among NLSH that best describes the subject of the work is recommended

⦿ Method & Process

- Using the table of contents and reviews of online bookstores
- Developing the machine learning model based on an automatic classification algorithm, AttentionMeSH

Project 1



Online bookstore'
data

주제어 추천

주제어 추천

서지 질의응답

민원 질의응답

ISBN

ISBN 9788965749202

가져오기

상행하기

작업 하기

초기화

ISBN

9788965749202

표제

태백산맥 세트(등단 50주년)

저자

조정래

소개글

1948년 10월, 전남 여수에 주둔하고 있던 국군 14연대 군인들이 반란을 일으킨다. 반란을 진압하기 위해 파견된 미군 1사단은 여수와 순천이 그들 손에 넘어가고, 남부지방은 중심으로 한 민간 학살지역이 발표될 전망이다. 그들은 인민재판에 임하여 자신들을 비롯한 이른바 반동세력을 공개처형한다. 하지만 도발적인 반란군 진압작전에 밀려 반란군은 산악지역으로 퇴각하고, 발표를 장악했던 민정청도 안창민, 하대치 등과 함께 임시, 불치산 두령에 돌입했다. 그 즈음 1948년 10월, 남부지방의 민정청이 발표된 미군 수적 지원에 따라

북차

제1부 항의의 모닥불 1권 1.얼굴 없는 새벽 | 2.가슴으로 이어진 물줄기 | 3.인족의 발견 | 4.소화, 하얀 꽃이라는 이름의 무당 | 5.조계산 속악 | 6.나라가 공산당 행글과 지주가 빨갱이 행군당해요 | 7.그리고 청년단 | 8.이년 이전의 인간 | 9.문답이 가시내, 불자도 참 함하게 변했다 | 10.광악(砲臺) 2권 11.세포 | 12.구만리창천을 떠도는 구름 | 13.냉철한 비탄을 생리로 가진 여사의 정체는 무엇인가 | 14.까마귀매 | 15.기습이다! | 16.갈매기는 먹을 수 있는 꽃 | 17.배고픈 몸과 웃음과 외간 | 18.소설 | 19.새가 창고에 그 똥자국을 새기지 못

DDC
DDC
국립중앙도서관 주제어
교보문고 주제어 한국소설 역사소설 대하소설

관한도 분단 전쟁

결과

표제 기반 결과

추천 결과가 없습니다.

results based on reviews

순위	항목	국립중앙도서관 주제어	학률
1	한국소설	한국 소설[韓國小說](KSH1998028419)	0.98322934
2	역사소설	역사 소설[歷史小說](KSH1998029910)	0.95551014
3	전쟁	전쟁[戰爭](KSH19980004373)	0.6774543

results based on TOC

순위	항목	국립중앙도서관 주제어	학률
1	역사소설	역사 소설[歷史小說](KSH1998029910)	0.9942926
2	한국소설	한국 소설[韓國小說](KSH1998028419)	0.92302704
3	대하소설	대하 소설[大河小說](KSH2000000137)	0.9157995

TOC, Review data
from a specific ISBN

Korean fiction

Historical fiction

Project 1

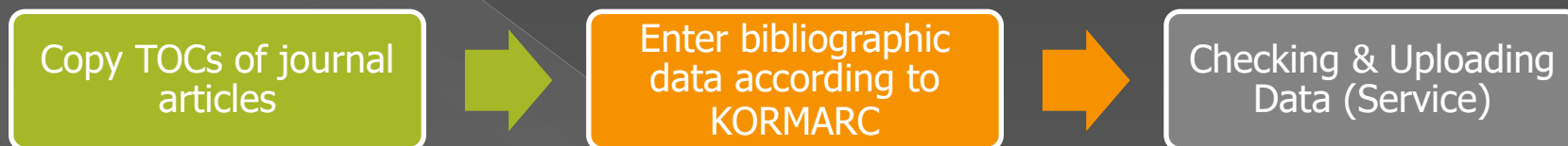
● Results & implications

- When using the table of contents, the recommended terms are more accurate than the introductory text
- The higher the number of assignments or the more frequent subject headings appear in the table of contents, the higher the accuracy
 - It must be mentioned at least 100 times in the training data for meaningful results
- Therefore, the trendy or new terms are not recommended well or their accuracy is low

Project 2

Workflow

- > Whole process is done manually by one staff (5,000 per year)



001		KS1000967846
005		20210730092521
008		210722s2021 bnk 000 eng
040		▼a011001
052	02	▼a615.05▼bk84▼c28(2)
245	00	▼aCommon data model-based real-world data for practical clinical practice guidelines : ▼bclinical pharmacology perspectives / ▼dYoomin Jeon, ▼eYoona Choi, ▼eEsther Hehsun Kim, ▼eSeonYeong Oh, ▼eHoward Lee
300		▼ap. 67-72 ; ▼c28 cm
545		▼aYoomin Jeon, Department of Transdisciplinary Studies, Graduate School of Convergence Science and Technology, Seoul National University ; Center for Convergence Approaches in Drug Development, Graduate School of Convergence Science and Technology, Seoul National University
545		▼aYoona Choi, Department of Transdisciplinary Studies, Graduate School of Convergence Science and Technology, Seoul National University ; Center for Convergence Approaches in Drug Development, Graduate School of Convergence Science and Technology, Seoul National University

title, author, affiliation, extent, keywords, journal information, etc.

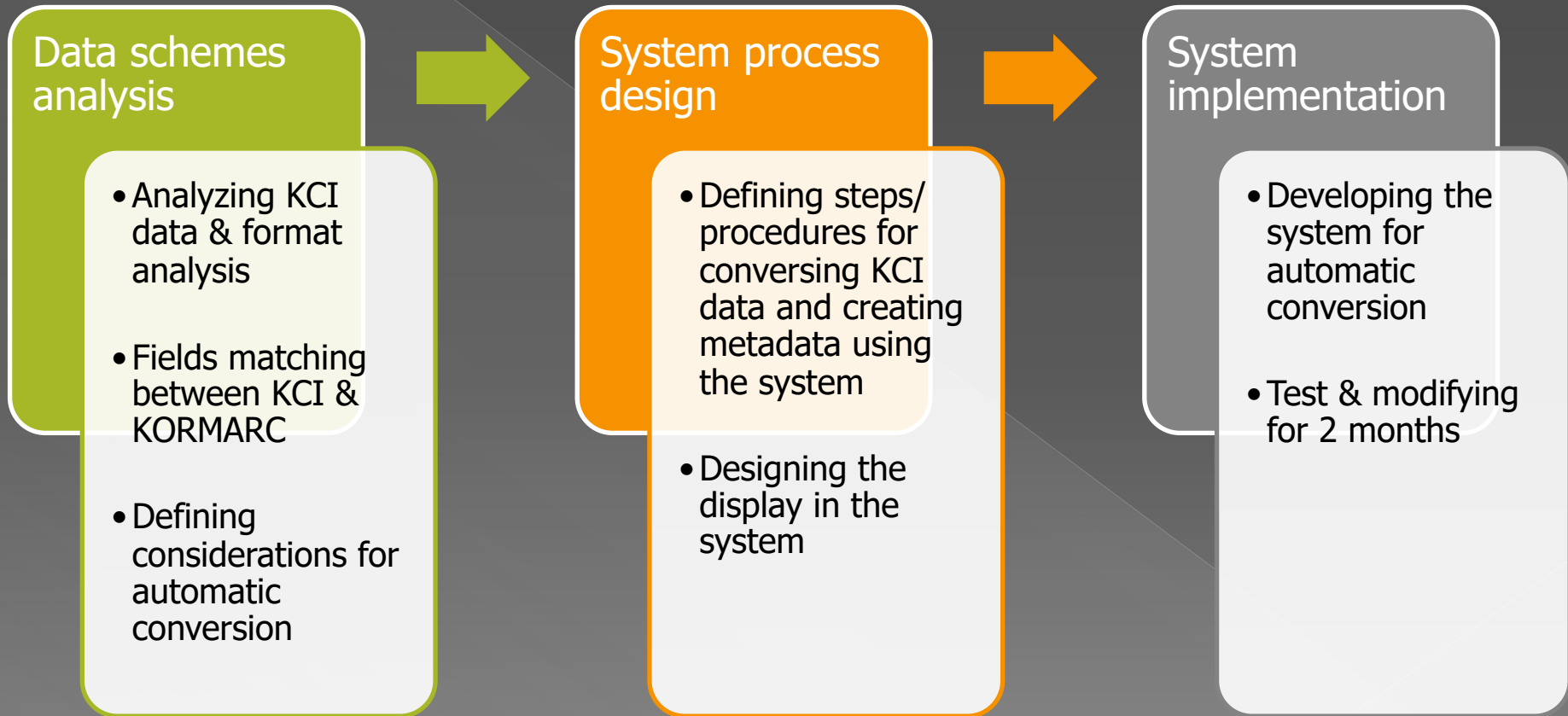
Project 2

◎ Purpose

- Automatic conversion of KCI to KORMARC data to improve work efficiency and ease of use based on the Korean Journal of Citation Index (KCI) data
 - KCI is a bibliographic database of domestic academic journals operated by the National Research Foundation of Korea
 - Department of Collection Development acquires and uses new data every month in Excel format

Project 2

Method & Process



Project 2

● Results & Implications

- Auto conversion evaluation
 - Highly auto-converted : 12 elements
 - Partially converted : 6 elements
 - Low level (data not included) : 8 elements
- Overall, there is few problem in identifying and searching the articles with auto-converted data without data correction or additional description
- We plan to fully apply auto-conversion for article metadata from next year

Q&A

- ◎ CHOI, Yoon Kyung yoonykchoi@korea.kr