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
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
### 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

TOC, Review data from a specific ISBN

Korean fiction

Historical fiction
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)

  Copy TOCs of journal articles → Enter bibliographic data according to KORMARC → Checking & Uploading Data (Service)

  
  | 001 | KSI000967846 |
  | 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 \(▼b\) clinical pharmacology perspectives ▼▼y ▼aYoomin Jeon, ▼aYoona Choi, ▼aEsther Hehsun Kim, ▼aSeon Yeong Oh, ▼aHoward Lee |
  | 300 | ▼a ap. 67-72 ▼b ▼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**

**Data schemes analysis**
- Analyzing KCI data & format analysis
- Fields matching between KCI & KORMARC
- Defining considerations for automatic conversion

**System process design**
- Defining steps/procedures for conversing KCI data and creating metadata using the system
- Designing the display in the system

**System implementation**
- Developing the system for automatic conversion
- Test & modifying for 2 months
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 yoonkchoi@korea.kr