https://doi.org/10.23106/dcmi.952653604

DCPAPERS

DCMI 2026

Proceedings of the International Conference on Dublin Core and Metadata Applications 2026

ISSN: 1939-1366

Exploring How Korean-Language Speakers Search for Korean-Language Materials and the Potential Role of AI in Cataloguing Materials in Non-Roman Scripts

Rinna L Kim [1]

[1] University of Toronto

Abstract

This exploratory study examines how Korean-language speakers search for Korean-language materials using an Online Public Access Catalogue (OPAC) at a public library in the Greater Toronto Area (GTA), Canada. Using a think-aloud protocol and semi-structured interviews, the study finds that participants encounter difficulties due to insufficient and inaccurate Korean-language metadata and inconsistent, unintuitive Romanization practices in the catalogue. The study highlights the need to improve cataloguing standards and practices to support the inclusion of accurate metadata in non-Roman scripts and the adoption of more user-friendly Romanization practices to enhance access. The study also suggests that emerging AI technologies have the potential to support the development of more equitable and accessible library catalogues.

1. Introduction

In the North American context, catalogues of library materials in non-Roman languages often lack metadata in the original languages and scripts. Lee [1] reported that WorldCat contains 52 million records in 67 non-Roman languages and that "on average, 30 percent of them lack their original script" (para. 5). Lee and Choi [2] further discussed whether Korean, Chinese, and Japanese metadata were included for the Harry Potter series or the Bible across selected public library websites in the United States. The study revealed that "half of the selected websites provided some basic non-English descriptions from the item pages," while "none provided non-English subject information" (p. 445).

In addition, catalogue records commonly rely on Romanized metadata rather than original scripts. This practice limits user access, as Romanized forms are often difficult for speakers of non-Roman languages to understand [3]. Nomura and Caidi [4] found that Japanese mothers in Toronto had difficulty finding Japanese-language materials due to challenges in Romanizing their own language.

The absence of metadata in non-Roman languages and scripts, and the reliance on Romanization, create significant search barriers, making it difficult for users to find the materials they need. To enhance user access, the inclusion of metadata in original languages and scripts is therefore critical, as it supports representation and promotes linguistic justice. Given the large number of non-Roman-language materials in library catalogues, emerging AI may support catalogue development despite limited human resources or time. To ensure such AI applications are meaning-driven and equitable, it

Student Forum

Published Sep 10, 2026

Correspondence to
Rinna L Kim
rinna.kim@mail.utoronto.ca

OPEN ACCESS

Copyright © 2026 Kim. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which enables reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator.

is essential to understand which user needs should be prioritized and which challenges should be addressed first.

As part of this broader discussion, this exploratory study examines how Korean-language speakers search for Korean-language materials at a public library in the Greater Toronto Area (GTA), Canada. The study aims to understand how, and to what extent, current metadata hinders users from searching for Korean-language materials, and what improvements are necessary to ensure equitable access.

2. Methodology

The study employs a think-aloud protocol and semi-structured interviews. Five Korean-language speakers aged 18 and older, living in the GTA, were recruited to complete three search tasks using the online catalogue of a public library in the GTA. The tasks include searching for (1) the Korean translation of the book *Crying in H Mart*, (2) nine Korean movies directed by Bong Joon Ho, and (3) the Korean film *Myeongnyang*.

All participants identified Korean as either their mother tongues and/or the most used languages at home, or reported at least full professional proficiency in Korean. In addition, they were non-experts in Online Public Access Catalogues (OPACs) or the Korean language, meaning that they had no formal study or work experience in these fields.

During the think-aloud sessions, participants were asked to verbalize their thoughts as they searched. After the think-aloud sessions, semi-structured interviews were conducted. The interview questions focused on why participants chose specific search terms or strategies. For data analysis, inductive thematic analysis was used, sensitized by Kuhlthau's Information Search Process (ISP) model.

3. Relation to Kuhlthau's ISP Model

This study employs Kuhlthau's ISP model as a sensitizing concept in inductive thematic analysis. Each stage of the model was conceptualized in relation to the study and used to understand participants' thoughts, actions, and feelings. The following sections outline the key findings of the study and discuss the challenges experienced by the participants, and factors that may have contributed to these challenges.

4. Findings

The study found that Korean-language speakers faced challenges during the search process. Some participants attempted to search in Korean but were unable to retrieve results when Korean metadata was absent. Others expressed surprise when an unexpected Romanized form of Bong Joon Ho's name, specifically Pong, Chun-ho (Romanized using the McCune-Reischauer system) appeared in the search results. In addition, one participant did not recognize McCune-Reischauer Romanization in the search results, which led to confusion. The participant subsequently used Naver (a Korean search engine) to interpret a Romanized Korean film title that they could not recognize.

In some cases, participants failed to retrieve results even when Korean metadata was present. Although they entered the correct Korean syllables of a book title, they were unable to retrieve the book, as the metadata contained incorrect word spacing (e.g., between a noun and a particle), which does not reflect standard Korean writing conventions. Others experienced confusion due to incomplete search results resulting from inconsistent Romanization of a personal name. For example, Korean films directed by Bong Joon Ho were catalogued under different spellings, such as Pong, Chun-ho and Bong Joon Ho.

5. Factors Contributing to these Challenges

These search challenges appear to stem from cataloguing standards that do not mandate the inclusion of Korean, the reliance on Romanized Korean forms that are unintuitive for Korean-language speakers, and guidelines that recommend practices inconsistent with Korean grammar conventions. Specifically, Fletcher and Toves [5] note that, depending on the context, the *Resource Description and Access* (RDA) standard, which is widely used for descriptive cataloguing, permits the use of Romanized metadata instead of non-Roman scripts for materials written in those languages. In addition, the McCune-Reischauer system, which is commonly used to Romanize Korean in North American libraries, does not align with how many Korean-language speakers perceive or Romanize the language [6]. The *Program for Cooperative Cataloguing* (PCC) guidelines further recommend word spacing practices in Korean metadata based on the McCune-Reischauer system [7], which are inconsistent with Korean writing conventions.

6. Conclusion

This study examined how Korean-language speakers search for Korean-language materials. The findings revealed that Korean-language speakers often experience search challenges due to the exclusion of Korean, unintuitive Romanization practices, and inconsistent and inaccurate metadata. These challenges appear to stem from standards and guidelines that create barriers to multilingual access, particularly for non-Roman language materials, including Korean-language materials. This study contributes to understanding how current metadata for non-Roman script materials can hinder access. It also highlights the need for improvements in cataloguing standards and practices for equitable access, specifically the inclusion of accurate Korean-language metadata and more user-friendly Romanized Korean metadata in catalogues of Korean-language materials.

7. Potential Role of AI in Cataloguing Materials in Non-Roman Scripts

AI may offer practical support in cataloguing multilingual materials. Research has demonstrated the potential of using AI to catalogue materials written in non-Roman scripts. Gamage and Wanigasooriya [8] discussed whether Generative AI is effective in creating Machine-Readable Cataloguing (MARC) records using book covers, title

pages or copyright pages. While the results indicated that AI may be less reliable when processing "local language texts," such as Sinhala and Tamil (p. 240), the cataloguing of five multilingual samples, including Arabic, Chinese, French, Korean and Spanish was successful, although difficulties were observed due to Chinese-language pages written in the top to bottom format.

Additionally, in response to budget cuts in libraries and difficulties in hiring librarians proficient in specific languages and with cultural expertise, Deng and Zhu [9] developed an AI system for generating Faceted Application of Subject Terminology (FAST) headings for Chinese materials. Although the study did not emphasize the importance of representing the languages used in the materials, the findings demonstrated that AI has potential as an assistant for subject heading assignment, suggesting it can process non-Roman languages.

Sun [10] also examined the use of AI to convert titles Romanized in Wade-Giles, the former standardized system for Romanizing Mandarin Chinese, into Pinyin, the current and widely used Romanization system. The findings showed that although different AI systems and inputs yielded varying results and there were some inaccuracies and inconsistencies, AI can be useful for Romanization conversion, helping librarians reduce the time for cataloguing.

Based on these studies, AI could assist in cataloguing Korean materials by potentially (1) recognizing and processing Korean text on title pages or book covers and extracting accurate bibliographic information in Korean, and (2) converting the McCune-Reischauer Romanization into a more user-friendly Romanization system. These forms of assistance may support libraries, librarians, and cataloguers in improving cataloguing to support equitable access, given the challenges of maintaining expertise in specific languages, scripts, and cultural contexts and the scale of non-Roman script materials in the North American library context.

Acknowledgements

This paper is based on the author's master's thesis at the University of Toronto. I would like to express my sincere gratitude to Dr. Silvia Vong for their invaluable guidance and support throughout my thesis research.

References

[1] C. Lee, "UCLA library/OCLC non-Roman script project bridges access gaps to research materials." [Online]. Available: https://www.arl.org/blog/ucla-library-oclc-non-roman-script-project-bridges-access-gaps-to-research-materials/

[2] T. H. Lee and I. Choi, "Multilingual support in digital environments: A case study of online catalogs in US public libraries, with a focus on Korean-language speakers," *The Library Quarterly*, pp. 437–454, 2023, doi: 10.1086/726321.

[3] H. H. Wellisch, "Bibliographic access to multicultural collections," *Library Trends*, pp. 223–244, 1980, [Online]. Available: https://hdl.handle.net/2142/7136

[4] T. Nomura and N. Caidi, "Heritage language acquisition and maintenance: Home literacy practices of Japanese-speaking families in Canada," *Information Research*, 2013, [Online]. Available: https://informationr.net/ir/18-3/colis/paperC37.html

[5] P. V. Fletcher and J. Toves, "An automated Cyrillic script project to increase non-Latin access in the catalog: Notes, observations, and conclusions," *Cataloging & Classification Quarterly*, pp. 346–357, 2023, doi: 10.1080/01639374.2023.2229823.

[6] S. Kim, "Romanization in cataloging of Korean materials," *Cataloging & Classification Quarterly*, pp. 53–76, 2006, doi: 10.1300/J104v43n02_05.

[7] Library of Congress, "PCC guidelines for creating bibliographic records in multiple character sets," 2017. [Online]. Available: https://www.loc.gov/aba/pcc/bibco/documents/PCCNonLatinGuidelines.pdf

[8] R. C. G. Gamage and P. Wanigasooriya, "Using Generative AI for bibliographic description: A study with ChatGPT 4," *Journal of the University Librarians Association of Sri Lanka*, pp. 227–256, 2024, doi: 10.4038/jula.v27i2.8083.

[9] S. Deng and L. Zhu, "Exploring the future of library cataloging with AI and multilingual embeddings." [Online]. Available: https://digitalorientalist.com/2025/05/09/exploring-the-future-of-library-cataloging-with-ai-and-multilingual-embeddings/

[10] L. Sun, "Enhancing cataloging with Generative AI: Converting Wade-Giles to Pinyin," *Cataloging & Classification Quarterly*, pp. 267–283, 2025, doi: 10.1080/01639374.2025.2508956.