

Abstract No: MO-23

Implementing an Automated Personal Name Authority File for Enhanced Information Retrieval in Sri Lankan Library Online Public Access Catalogues

P. L. Wanigasooriya^{*1} and R. W. P. De Silva²

^{*1}Department of Library and Information Science, University of Kelaniya

²Technical Lead, Elliott moss consulting, Singapore

priyanwada@kln.ac.lk*

In the digital era, managing vast information collections while ensuring seamless access for patrons is a challenge faced by Sri Lankan libraries. A critical aspect that significantly affects information retrieval is the accuracy and consistency of personal names in Online Public Access Catalogues (OPACs). To address this issue, this paper proposes the implementation of an Automated Personal Name Authority File (APNAF) system, which offers a viable solution to enhance information retrieval, improve user experience, and foster an organized and user-friendly OPAC environment. The objectives of implementing APNAF in Sri Lankan libraries encompass enhancing search precision and recall in OPACs, minimizing name-based retrieval errors and confusion, facilitating efficient data maintenance and updates, and promoting standardization and consistency in personal name entries. To conduct the study, an author's name dataset was selected from the National Library of Sri Lanka, one of the largest libraries in the country. The approach utilizes the Jaro-Winkler algorithm to analyze similarities in names, addressing complexities such as variations in names, spelling errors, and differences in word order. The dataset initially contained 77,000 records, which were subsequently refined to 44,000 unique data points. Through a thorough and meticulous examination, we discovered essential patterns, trends, and correlations within the dataset, providing valuable insights that greatly enriched our understanding of the subject matter. The study revealed pronounced similarities in names within multilingual datasets, accompanied by instances of confusion in transliteration. Clear cross-linguistic correlations came to light, notably in cases where Sinhala and English names exhibited shared linguistic components. These shared elements resulted in both visual and phonetic resemblances. The research delineated four distinct iterations of Sinhalese name clusters, which are summarized as follows: name similarity within multilingual datasets, transliteration conventions, cross-linguistic connections, and variations, along with exceptions. Further, our analysis revealed a trade-off between accuracy and recall in duplicate name detection and the Jaro-Winkler algorithm proved effective in identifying despite variations in spelling, typos, or minor differences in naming conventions. To address these challenges, the implementation of the system in Sri Lankan libraries can significantly enhance the efficiency of information retrieval. This system simplifies the process of entering data into the Online Public Access Catalog (OPAC) by facilitating the selection of unique author names for catalog records.

Keywords: OPAC, National Library of Sri Lanka, Personal Name Authority File, Jaro-Winkler algorithm, Library