Strategic Thesis Collection Management through Statistical Analysis

N. M. Karannagoda[1], A. K. K. K. Athukorala[2], K. A. C. P. Abeygunawardena[3], D. D. G. S. Ariyawansa[4]

[1][3][4]Main Library, University of Colombo, Sri Lanka

[2]Horizon Campus, Sri Lanka

The thesis collection at the University of Colombo's (UOC's) Main Library (ML) plays a vital role in undergraduate and postgraduate education, serving as a repository of scholarly work produced by the students in Masters, MPhil, and PhD degree programs in five faculties. The collection reflects the academic output of the UOC student and serves as a valuable resource for researchers, academic staff, and students. A comprehensive understanding of the Main Library Thesis Collection (MLTC) content and patterns is required for effective management and utilisation. Therefore, in this study, content analysis of the MLTC was done through statistical analysis. Statistical analysis of the thesis collections at the UOC Main Library is the main objective of this research. Further, the paper examined the (1) faculty-wise and year-wise publication of the theses, (2) degree level-wise distribution of the theses, and (3) subject-wise distribution of the theses within the MLTC. For this study, it has used all theses contained in the MLTC, and it is around 1827 theses. Data collection and analysis were done from July 2023 to September 2023. Excel data analysis tool pack and its features were used for data analysis. The MLTC contains theses from six faculties (Arts, Education (FED), Graduate Studies, Law, Management, and Science). According to the findings, FED is the most prominent faculty contributing to the library's thesis collection. The year-wise distribution of the theses at the MLTC indicates that the highest number is from 2000-2009, followed by 2010-2019. The collection includes 181 PhD, 280 MPhil, and 1366 master's theses. When considering the main ten categorisations of the Dewey Decimal Classification, 67% of these are in Social Sciences, and 12% are in Technology. Considering the physical condition of the theses, most of the theses in the MLTC are in excellent condition. It is evident the institution's commitment to maintaining the highest preservation standards. As a result of the findings, the following recommendations were made regarding the MLTC: (1) Increasing the contribution of the Faculty of Law, (2) Increasing the presence of PhD theses, (3) Enhancing representation in underrepresented subject areas, and (4) Digital preservation and access.

Keywords: Bibliometric Analysis, Library, Statistical Analysis, Thesis Collection, UOC

Background and Rationale

The thesis collections at the University of Colombo (UOC) Main Library play a vital role in undergraduate and postgraduate education, serving as repositories of scholarly work produced by students. The collections reflect the academic output of the UOC and serve as valuable resources for researchers, faculty, and students. The Main Library Thesis Collection (MLTC) spans six faculties (Arts (FAR), Education (FED), Management and Finance (FMF), Graduate Studies (FGS), Law (FLA) and Science (FSC)), making them a vital part of the scholarly community. This collection is located on the 4th floor of the library building in a special section called the Ceylon room. Annually, more than 15,000 library users register at the Main Library, which will require an enriched thesis collection.

Assessing the characteristics and patterns of the MLTC is crucial for several reasons. Understanding the composition and trends within the thesis collection provides insights into the research output and scholarly interests of the UOC's five faculties. Knowledge of such matters is helpful to academics, administrators, librarians, and faculty members who develop collections and allocate resources. To improve the relevance and comprehensiveness of UOC's thesis collections, UOC should identify areas of strength and areas that need further attention.

A comprehensive understanding of the content and usage patterns of the MLTC is required for effective management and utilisation. This task can be accomplished by doing through a statistical analysis, a quantitative approach. The statistical analysis of thesis collections provides critical insights into publication trends and subject distribution. Research findings can inform collection development, preservation efforts, and research promotion decisions.

However, up to now, a statistical analysis has not been directly focused on the MLTC, and this study will fill that gap. Therefore, a comprehensive study of library thesis collection based on descriptive statistics is needed.

The study is motivated by its potential impact on the management and utilisation of library thesis collection. Statistical analyses can provide insights into collection publication trends and subject distributions. Researchers, librarians, academics and administrators can use this knowledge to decide how to allocate resources, develop collections, and collaborate on research.

Based on the potential benefits of conducting statistical analyses of library thesis collection, this research paper examines (1) Faculty-wise and year-wise publication of the thesis, (2) Degree level-wise distribution of the thesis, and (3) Subject distribution and physical condition of the thesis within the library thesis

collection. These objectives aim to improve the understanding of the library's thesis collection and its potential as a resource for scholarly research.

Research objectives

Main Objective

To do a comprehensive analysis on content (content analysis) and usage patterns of the thesis collections of the University of Colombo, Main Library.

Specific Objectives

1. To examine the faculty-wise and year-wise publication of the theses

Identifying the distribution of theses across years and faculties is the first objective of the research. Additionally, the collection's growth rate will be analysed over time, and any significant shifts or changes in publication trends will be determined.

2. Analyse the degree level-wise distribution of the theses

Under the second specific objective, it will analyse the thesis collection by the degree program level (Master, MPhil, and PhD.)

3. To examine the subject distribution and physical condition of the theses within the library thesis collection

Identifying the dominant subject areas and disciplines represented in the collection will be the third objective of the study. Further, it will intend to analyse the physical condition of the theses within the MLTC.

Significance of the Study

Researchers and academic institutions can benefit from this research, which could contribute to library and information studies. This study provides valuable insights and benefits in the following areas by conducting a comprehensive statistical analysis of the library thesis collection:

1. Advancement of Knowledge

This study contributes to existing knowledge by analysing library thesis collections statistically. There is a notable lack of studies examining thesis collections' unique characteristics and patterns.

2. Evidence-Based Collection Development

As a result of this study, academics, administrators, librarians, and collection development professionals can make informed decisions regarding resource allocation and enhancement of the thesis collection. Identifying faculty-wise, year-wise, and degree-level-wise theses distribution and subject distribution within the library thesis collection will provide valuable insights into its strengths and weaknesses. Align the library thesis collection with the evolving needs of the academic community, and this evidence-based approach can guide the acquisition of new theses, preservation efforts, and targeted promotion of specific research areas.

3. Enhanced Accessibility and Discoverability

This research aims to enhance the accessibility and discoverability of library thesis collections by statistically analysing them. Accessibility and discoverability improve the visibility and impact of theses and the institution's research output.

Literature Review

Statistical analysis is a quantitative method commonly used to analyse data. Although statistical analysis has been extensively applied to journals, conference proceedings, and other scholarly publications, its application to library thesis collections remains relatively limited.

A notable study conducted at the State University of Medan, Nigeria, by Sayekti et al. (2022) examined statistical indicators of a university thesis collection. In addition, it provided insights into authorship patterns and publication trends. According to their results, submitting theses has increased significantly, with a concentration of authors from specific departments (Sayekti, et al., 2022).

Rosenberg (2015) uses statistical analysis to analyse the Master's and Doctoral theses collections of the Department of Sociology and Anthropology at the Hebrew University of Jerusalem, Israel. A total of five variables were identified and quantified in this study: type of material (monographs, journals, or others), format (print, electronic, or both), language of material (English, Hebrew, or other), publication year, publication age, and availability of items at Hebrew University (Rosenberg, 2015).

Burnett & Lingam (2014) conducted another essential study at the University of the South Pacific in Fiji using statistical analysis to analyse their thesis collection. A study was conducted at the University of South Pacific to discover trends and patterns in postgraduate research (Burnett & Lingam, 2012).

Huang & Lin (2012) conducted another study in Taiwan on the citation analysis of theses in a specific field. As a result, influential theses and citation patterns were revealed within the collection. The findings of these studies demonstrate the potential of statistical analysis to uncover valuable insights into thesis collections (Huang & Lin, 2011).

Uçak & Al (2009) also conducted a fascinating study in Germany using statistical analysis to find the differences between 391 theses between 1968 and 2007. Moreover, the study found that literature obsolescence, the language of resources, journal or monograph usage, citations, and authors vary by discipline (Uçak & Al , 2009).

Clarke (2000) conducted a statistical analysis of the subject distribution in the University of the West Indies, St. Augustine Campus thesis collection. In his research, he identified dominant subject areas and revealed trends in the representation of different disciplines, revealing the intellectual breadth of the theses (Clarke, 2000).

Methodology

This section describes the methodology for analysing the library's thesis collection. The study examines the faculty-wise, year-wise, and degree level-wise theses distribution and subject distribution within the collection. Data was collected from all the theses within the MLTC from June 2023 to August 2023 and analysed.

Sandra Eldridge developed the "Onion Model" to guide research design and methodology. It has multiple layers representing various aspects of a study. Using Sandra's Onion Model, let's discuss the methods behind the survey of "statistical analysis of the library thesis collection".

Research Philosophy

This research study is based on quantitative analysis and focuses on collecting and analysing quantitative data to establish general patterns and trends; thus, it can define this study as a positivist research study.

Research Approach

In the deductive approach, the researcher will get information from previous research, like the selected topic, to conduct the present study. Therefore, in this research study, the research approach used by the researcher is a deductive one.

Methodological Choice

The methodological choice in this study is primarily quantitative. It involves systematically collecting and analysing numerical data from the library thesis collection.

Research Strategy

A statistical analysis strategy is used to examine the faculty-wise, year-wise, degree-level-wise thesis distribution and subject distribution within the collection. Data was collected manually from the library system by referring to each thesis individually. As a result, the research strategy can be considered a survey method.

Time Horizons

In this study, collecting the data from the sample was done within a single time frame. Therefore, the time horizon for this research can be identified as cross-sectional.

Data Collection

A comprehensive search was conducted to identify the library thesis collection for analysis. The collection includes Master's, MPhil, and Doctoral theses from six faculties (FAR, FED, FMF, FLA, FGS, and FSC). Bibliographic information about theses, including titles, authors, publication years, class numbers, and thesis level, was retrieved from the library's online catalogue. Faculties and class numbers were manually collected.

Data Cleaning and Preprocessing

The collected data underwent a rigorous cleaning process to ensure accuracy and consistency. The process involved removing duplicates, correcting errors, and standardising author names and publication years. For ensure the reliability of the results, any thesis with incomplete or missing data was excluded from the analysis. The thesis abstracts were preprocessed using stemming, stop-word removal, and keyword extraction techniques to enhance the accuracy of subject categorisation and topic analysis.

Statistical Analysis and Visualisation

Descriptive statistics, such as mean, median, and standard deviation, were calculated to summarise the quantitative characteristics of the dataset. Various visualisation techniques, including bar charts, pie charts and tables, were utilised to represent the analysis findings visually.

Results and Discussion

Faculty-wise and year-wise publication of thesis

The library thesis collection contains 1827 theses from six faculties (FAR, FED, FGS, FLA, FMF, and FSC).

Table 1: Distribution of theses by faculty at the library's research collection

Faculty	Number of thesis
Faculty of Arts	353
Faculty of Education	764
Faculty of Graduate Studies	364
Faculty of Law	44
Faculty of Management	215
Faculty of Science	87

Based on the data, it's evident that the FED produces significantly more theses than other faculties. This indicates that the FED is the most prominent faculty contributing to the library's thesis collection. Also, the FLA contributes the least to the library thesis collection, with a relatively low research output.

Year-wise distribution of the thesis located at the MLTC indicates that the highest number of theses are from 2000-2009, followed by 2010-2019. The least number of theses can be found from 1960-1969. According to the data, there has been a trend of increasing research output over the years, with fluctuations in some periods. The number of theses increased significantly from the 1970s to the 2000s and then stabilised in the 2010s. Due to the data collected early in the decade, the number of theses has sharply decreased in the 2020s.

Table 2: Distribution of thesis by year in the library's research collection

Publication Year	Number of Theis	Percentage %
1960 - 1969	1	0%
1970 - 1979	96	5%
1980 - 1989	253	14%
1990 -1999	322	18%
2000 - 2009	585	32%
2010 - 2019	548	30%
2020 - August 2023	21	1%

Degree level-wise distribution of the theses

There are numerous theses in the library's extensive research collection arranged according to the degree level of the academic achievement they represent. This stratification reveals the institution's commitment to fostering intellectual growth and rigorous scholarly exploration across educational milestones. The collection includes 181 theses awarded the prestigious Doctor of Philosophy (PhD) title. PhD theses exemplify the pioneering spirit of inquiry and innovation that typifies the PhD experience. The authors of each thesis contributed groundbreaking insights to their chosen field by pushing the boundaries of knowledge. The collection contains 280 Master of Philosophy theses that represent the culmination of diligent research and intellectual engagement. This type of thesis bridges the gap between advanced scholarship and original inquiry, demonstrating an author's mastery of a particular domain and commitment to contributing thought-provoking perspectives. A total of 1366 master's theses comprise the largest segment of the collection. Across various disciplines, these theses illustrate the rich intellectual tapestry of higher education. In addition to contributing to the existing body of knowledge, each Master's thesis represents the author's dedication to studying their area of study.

In classifying theses according to degree level, the library emphasises its commitment to capturing the multifaceted journey of scholarly development. As well as celebrating individual researchers' achievements, this arrangement also serves as a valuable resource for academics, students, and enthusiasts eager to explore various academic pursuits. At each stage of learning and discovery, the distribution of theses reflects the institution's commitment to nurturing academic excellence.

Table 3: Distribution of degree-level theses in the library's research collection

Degree Level	Number of Thesis
Doctor of Philosophy	181
Master of Philosophy	280
Master	1366

Subject distribution within the library thesis collection

According to the analysis, the library thesis collection comprises theses from the following subject areas. There are around 1227 theses from the social sciences discipline out of 1827 total theses in the collection. In other words, nearly 67% of the theses belong to the Social Sciences discipline. Around 15% of the whole thesis is attributed to the Technology discipline in the library thesis collection. Approximately 1% comes from History, Geography, Language, Literature, Philosophy and Psychology.

Table 4: Theses distribution by subject - Library research collection

Subject Area	Number of Thesis	Percentage %
Arts and recreation	12	1%
Computer science, information and general works	113	6%
History and geography	26	1%
Language	24	1%
Literature	22	1%
Philosophy and psychology	27	1%
Pure science	90	5%
Religion	11	1%
Social sciences	1227	67%
Technology	274	15%

Physically, most of the theses in the library thesis collection are in excellent condition, according to the research study. Physically, most of the theses (52%) in the library thesis collection are in perfect condition, according to the research study. It is evident from these theses that the institution is committed to maintaining the highest preservation standards. In addition to being immaculately cared for, these theses provide insight into the institution's proactive approach to safeguarding its intellectual heritage. In addition, 41% of the theses in the thesis collection are in good condition, and 4% are in satisfactory condition. However, 3% of these are in urgent need of repair. As part of the institution's commitment to safeguard its collection for future generations, these require immediate attention to halt further deterioration.

Table 5: The physical condition of the thesis at the library research collection

Physical condition of the Thesis	Number of Thesis	Percentage %
Excellent	950	52%
Good	750	41%
Satisfactory	67	4%
Need to repair	25	1%
Urgent to repair	35	2%

Conclusion and Recommendations

Conclusion

The library's thesis collection has revealed a multifaceted portrait of academic excellence and institutional commitment. Research output, degree levels, subject areas, and physical condition reflect an institution that values scholarly pursuits. The library's thesis collection demonstrates its role as a hub of intellectual exploration, fostering a vibrant academic environment. The distribution of theses by year of publication, degree level, subject area, and physical condition indicated an institution deeply interested in disseminating, preserving, and making accessible knowledge. Academic engagement at the institution has progressed over time, as evidenced by the growth in research output. Diversity in degree levels emphasises the inclusive nature of the school's education. Diversity in subject areas reflects the institution's commitment to interdisciplinary education. The physical condition of the theses demonstrates the library's dedication to the long-term sustainability of its scholarly resources and that meticulous care has been given to the thesis' physical condition.

Recommendations

Amplifying the Faculty of Law's Contribution: FLA theses are notably underrepresented in the library's thesis collection. It is advisable to proactively increase the inclusion of law-related theses in the FLA's collection to enhance scholarly diversity and enhance the FLA's academic presence. With this initiative, the FLA can expand its legal scholarship resources, foster a more comprehensive educational environment, and reinforce its prestigious academic position.

Enhancing the Presence of Ph.D. Theses: Compared to Masters and M.Phil. theses, the library thesis collection has a relatively small number of PhD theses. To address the disparity and foster a more

comprehensive scholarly landscape, it is necessary to take proactive measures to increase the number of PhD theses in the collection. As a result, the institution can enrich its academic repository and further enhance its research ecosystem.

Enhancing Representation in Underrepresented Subject Areas: The library's thesis collection lacks representation in particular subject areas. It is recommended that efforts be focused on enriching theses within underserved subject areas to address this gap and develop a more well-rounded academic repository. The institution can enhance its collection by actively seeking and acquiring theses from these disciplines, ensuring a balanced and comprehensive representation of diverse academic interests. Embracing this approach aligns with the institution's commitment to fostering an inclusive and holistic research environment.

Digital Preservation and Access: Although the physical preservation of theses is carefully managed, digitisation efforts could make them more accessible. In addition to aiding conservation, digitisation also facilitates remote access for researchers worldwide.

.

References

Burnett, G., & Lingam, G. I. (2012). Postgraduate research in Pacific education: Interpretivism and other trends. *Springer*, 221-233.

Clarke, R. (2000). Thesis Writing and Preservation at the University of the West Indies, St. Augustine, Trinidad: A Librarian's Input. *World Libraries*.

Gingras, Y. (2016). *Bibliometrics and Research Evaluation: Uses and Abuses.* The MIT Press.

Huang, M.-H., & Lin, C.-S. (2011). A Citation Analysis of Western Journals Cited in Taiwan's Library and Information Science and History Research Journals: From a Research Evaluation Perspective. *The Journal of Academic Librarianship*, 34-45.

Rosenberg, Z. (2015). Citation Analysis of M.A. Theses and Ph.D. Dissertations in Sociology and Anthropology: An Assessment of Library Resource Usage. *The Journal of Academic Librarianship*.

Sayekti, R., Aditya, M., Nurhayani, Simahate, T., Yusniah, & Devianty, R. (2022). An Assessment of Information Control: Understanding Library Service Quality from Users' Perspectives. *Evidence-Based Library and Information Practice*, 88-108.

Uçak, N. Ö., & Al , U. (2009). The Differences Among Disciplines in Scholarly Communication: *Libri*.