

AUTOMATIC ELECTRONIC THESIS AND DISSERTATION GUIDELINES VERIFICATION

A Data Mining Tool For Improved Compliance Against Postgraduate Guidelines

Mubanga Chibesa | Albertinah Mooka | Gift Muwele | Lwiime Shansonga | Lighton Phiri

DEPARTMENT OF LIBRARY AND INFORMATION SCIENCE

UNIVERSITY OF ZAMBIA



INTRODUCTION

The University of Zambia provides training for postgraduate students. A key aspect of this training involves the production of an ETD manuscript specifically for postgraduate students pursuing a mode B as well as a mode C Masters degree and PhD students.

The University, through the Directorate of Research Innovation and Development (DRID) provides guidelines [4] that the formatting and layout of ETD manuscripts must conform to. The process of verifying if submitted ETDS conform to the guidelines is manual, proving to be tedious and time consuming for examiners.

Therefore this project aims at addressing the above challenges through the implementation of a tool that will leverage natural language processing, specifically Document Layout Analysis (DLA) Techniques [1] to automate this process and flag off parts of manuscripts that do not conform to the guidelines as well as provide actionable feedback.

RESEARCH OBJECTIVES

1. To Identify the University of Zambia postgraduate guidelines that ETDs manuscripts should conform to.
2. To analyse the University of Zambia archived historical ETDs compliance to postgraduate guidelines.
3. To Investigate the challenges faced when checking for ETDs compliance to postgraduate guidelines.
4. To Design and implement a software tool that will utilise document layout analysis (DLA) techniques that will automatically flag off portions of manuscripts that do not conform to the institutional guidelines.

METHODOLOGY

Using a mixed methods approach, document analysis will be employed to understand the postgraduate guidelines stipulated in the “Regulations and Guidelines for Postgraduate Studies” guidelines document; content analysis will be used on randomly sampled ETDs in order to experimentally determine the extent of the problem and, finally, a DLA Natural Language Processing model[3] will be developed and evaluated using standard DLA metrics such as Structure Similarity Index and Intersection over Union[2]. This DLA pipeline is hinged on Artificial intelligence and Natural language processing(NLP). We will also everage an open source package called Deepdoctetion for the implementation of the software tool.

PRELIMINARY RESULTS & DISCUSSION

Figure 1. Images representing the Guidelines (Right) and Sample ETD script with parts not complying pointed out (Left).

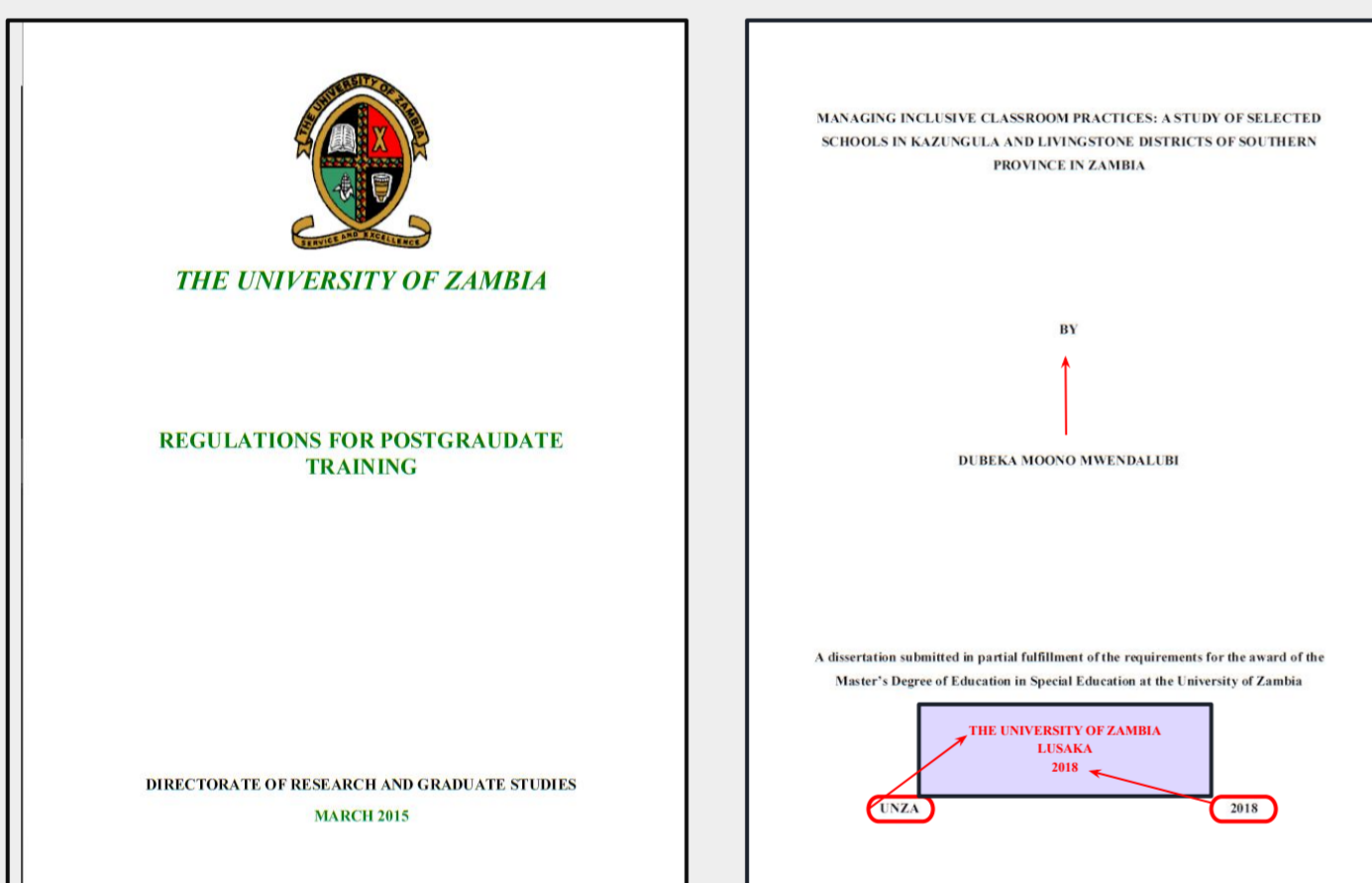


Figure 4 Below is a chart representing the average scores from the analysis of historical achieved ETD Manuscript of the entire schools at the University of Zambia on each preliminary section of the manuscript

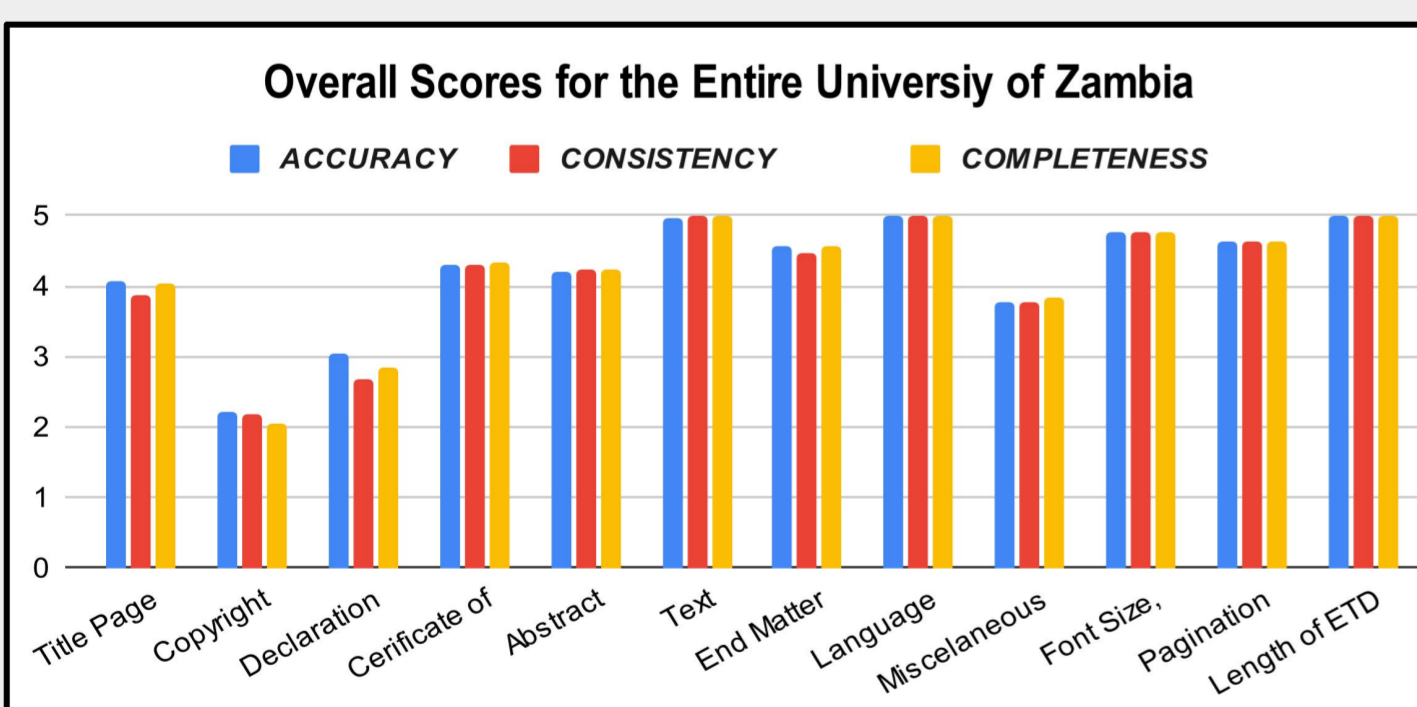


Figure 2. The chart below shows the responses gotten from Current Students on the challenges they faced when checking if their manuscript were in compliance with postgraduate guidelines

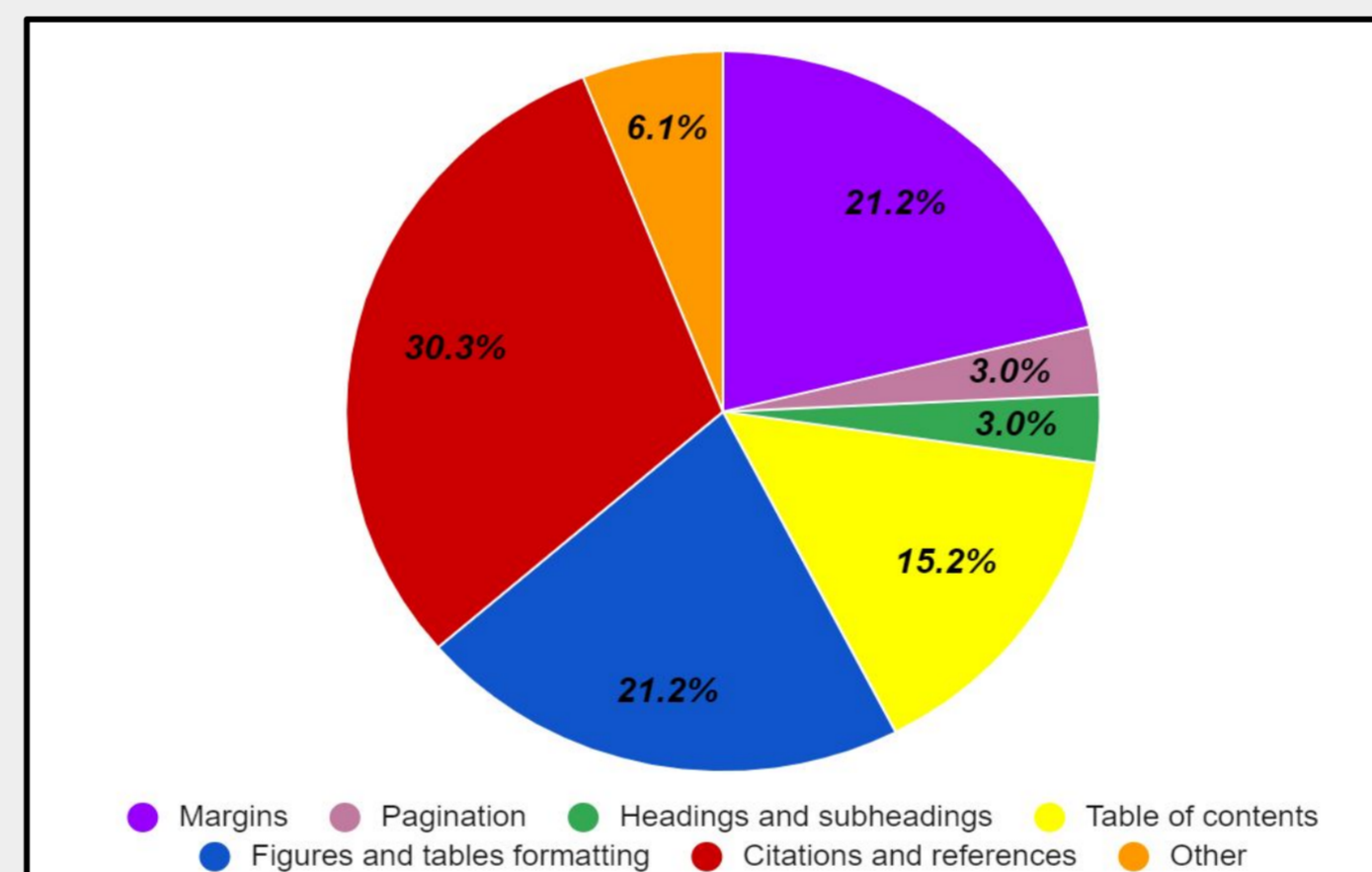


Figure 5 The diagram below illustrates how the software tool will operate once implemented.

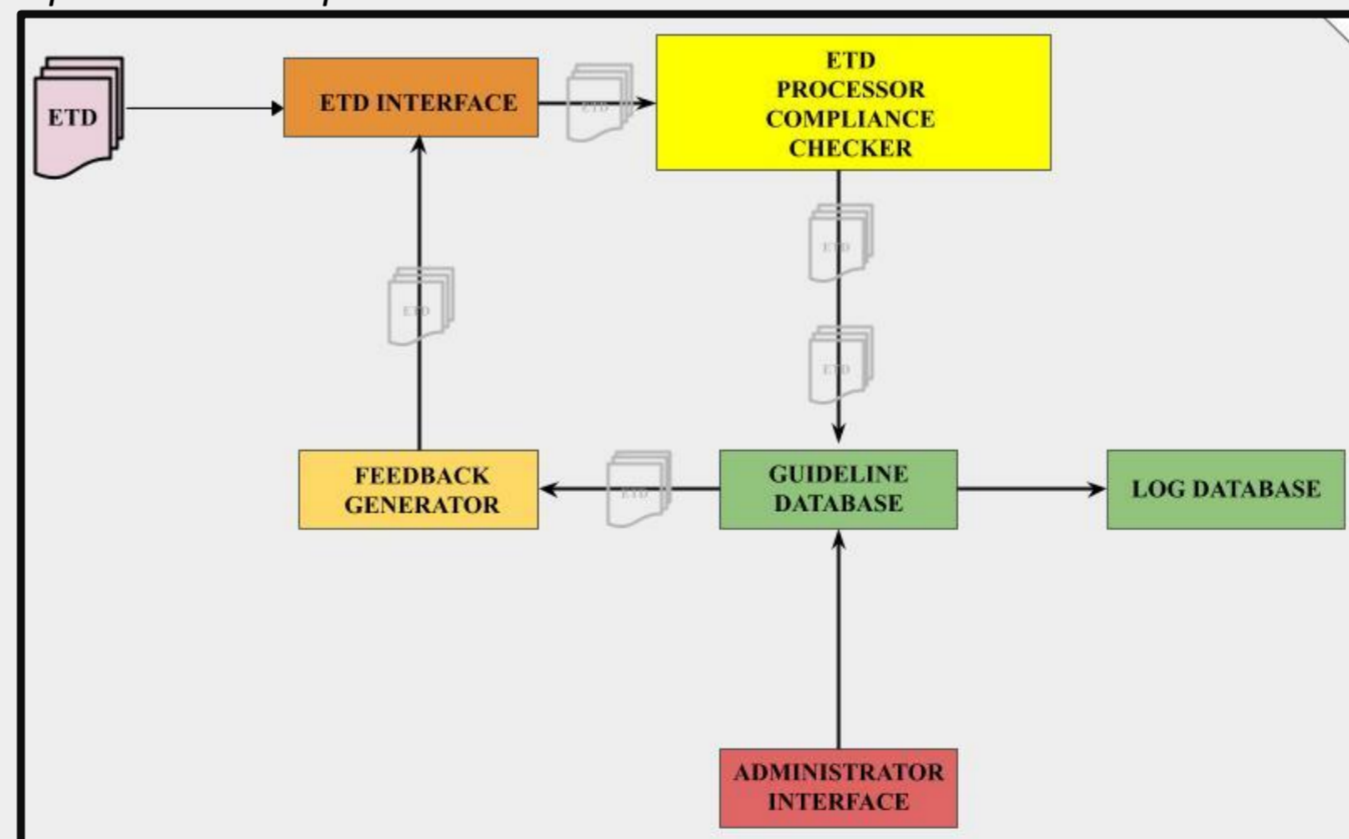


Figure 3. The diagram below show responses gotten from Alumni Students when asked How familiar they were with the DRID postgraduate guidelines for ETD formatting

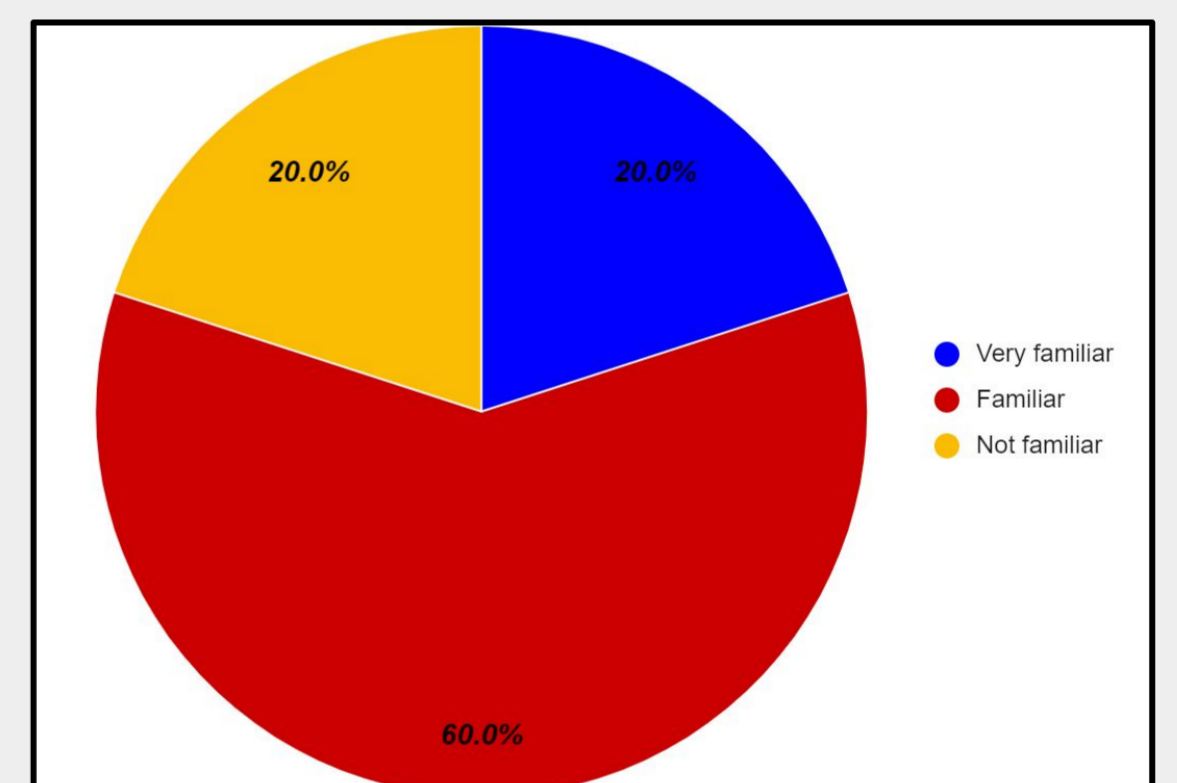
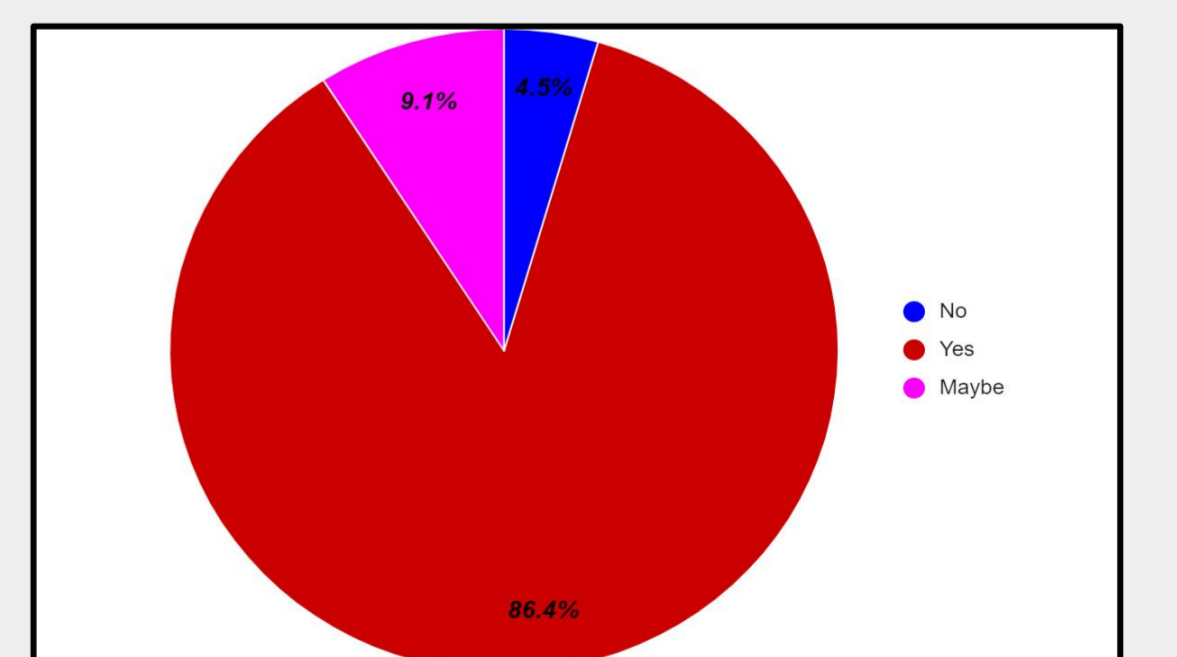


Figure 6 below show responses gotten from Current Students when asked if there is a need for an automated software tool to assist with checking if their manuscript conform to institutional guidelines?



CONCLUSION

The development of an ETD automatic guideline verification tool presents an opportunity to enhance efficiency as well as promote consistency in the quality of ETDs while alleviating the challenges faced in the process of manually checking for compliance consequently reducing the workload for students and examiners.

BIBLIOGRAPHY

- [1] Binmakhashen, G. M., & Mahmoud, S. A. (2019). *Document Layout Analysis: A Comprehensive Survey*. *ACM Comput. Surv.*, 52(6), 1–36.
- [2] Mishra, B. K., & Kumar, R. (2020). *Natural Language Processing in Artificial Intelligence*.
- [3] Razi, S., Glendinning, I., & Foltýnek, T. (2019). *Towards Consistency and Transparency in Academic Integrity*. Peter Lang GmbH, Internationaler Verlag Der Wissenschaften.
- [4] The University Of Zambia (2015). *Regulations For Postgraduate Training*.

DOI: https://graduate.unza.zm/images/files/pg_regulations.pdf

Project Team Members

- Chibesa Mubanga <2020031345@student.unza.zm>
- Mooka Albertinah <2020019299@student.unza.zm>
- Muwele Gift <2020014203@student.unza.zm>
- Shansonga Lwiime <2020002353@student.unza.zm>

Project Supervisor

- Dr.Lighton Phiri <lighton.phiri@unza.zm>

