

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

Higher Education Institutions worldwide enforce guidelines and academic approaches to ensure scholarly integrity and adherence to academic standards. The University of Zambia is not an exception. Just like most HEIs it offers training to postgraduate students and one of the key aspects of postgraduate training is producing an Electronic Thesis and Dissertation manuscript. The Directorate of Research Innovation and Development (DRID) at the University of Zambia provides guidelines [4] which stipulate how ETD's should be formatted. However, the process of checking for conformance is a manual and tedious procedure, resulting in submission of inconsistently formatted manuscripts in the Institutional Repository (IR).

To address this challenge our project seeks to implement a tool that will automate the process of checking ETD's compliance against established postgraduate guidelines. The tool will leverage data mining techniques to perform this task. More specifically, Document Layout Analysis (DLA)[1] will be the core approach used in the implementation. The tool will flag off portions of ETD manuscripts that do not conform to established guidelines. Hence, this will help resolve the inconsistencies in the format of submitted manuscripts.

RESEARCH OBJECTIVES

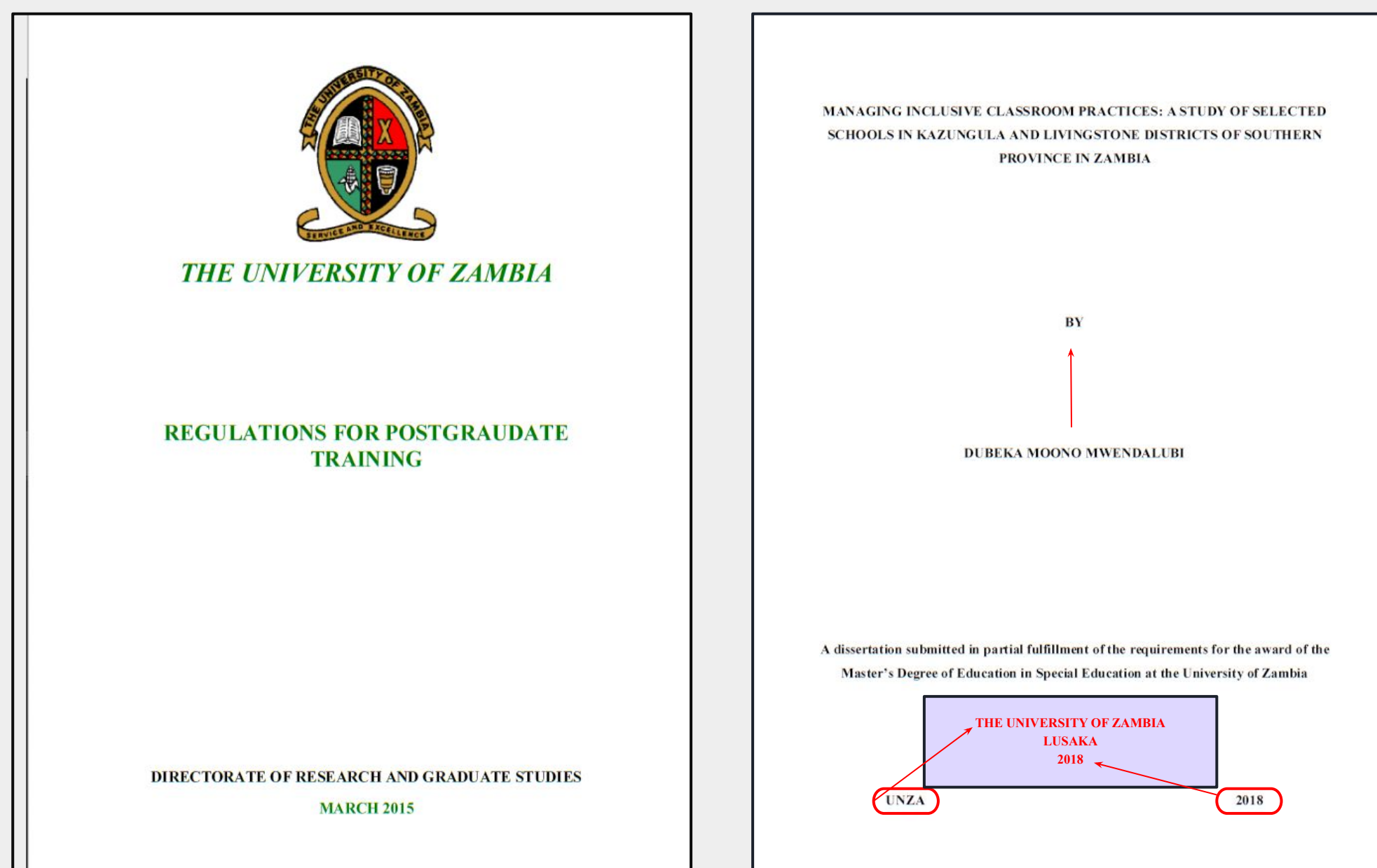
1. To Identify the University of Zambia postgraduate guidelines that ETDs manuscripts should conform to.
2. To analyse the University of Zambia archived historical ETDs compliance to postgraduate guidelines.
3. To Investigate the challenges faced when checking for ETDs compliance to postgraduate guidelines.
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

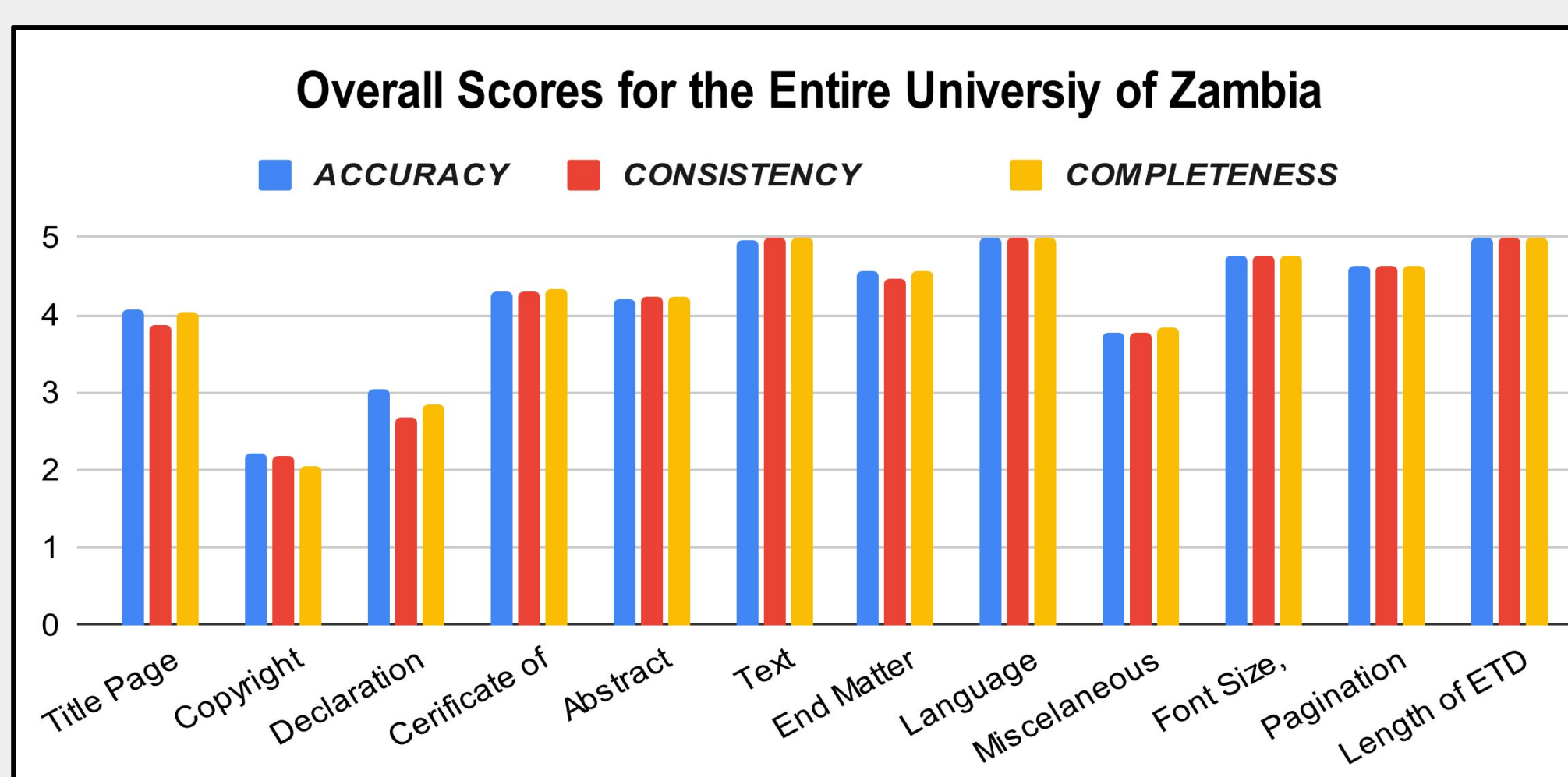
Using a mixed methods approach, document analysis will be employed to understand the postgraduate guidelines stipulated in the "Regulations and Guidelines for Postgraduate Studies" guidelines document; content analysis was used on randomly sampled ETDs in order to experimentally determine the extent 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 also leveraged an open source package called Deepdoctection for the implementation of the software tool.

RESULTS & DISCUSSION

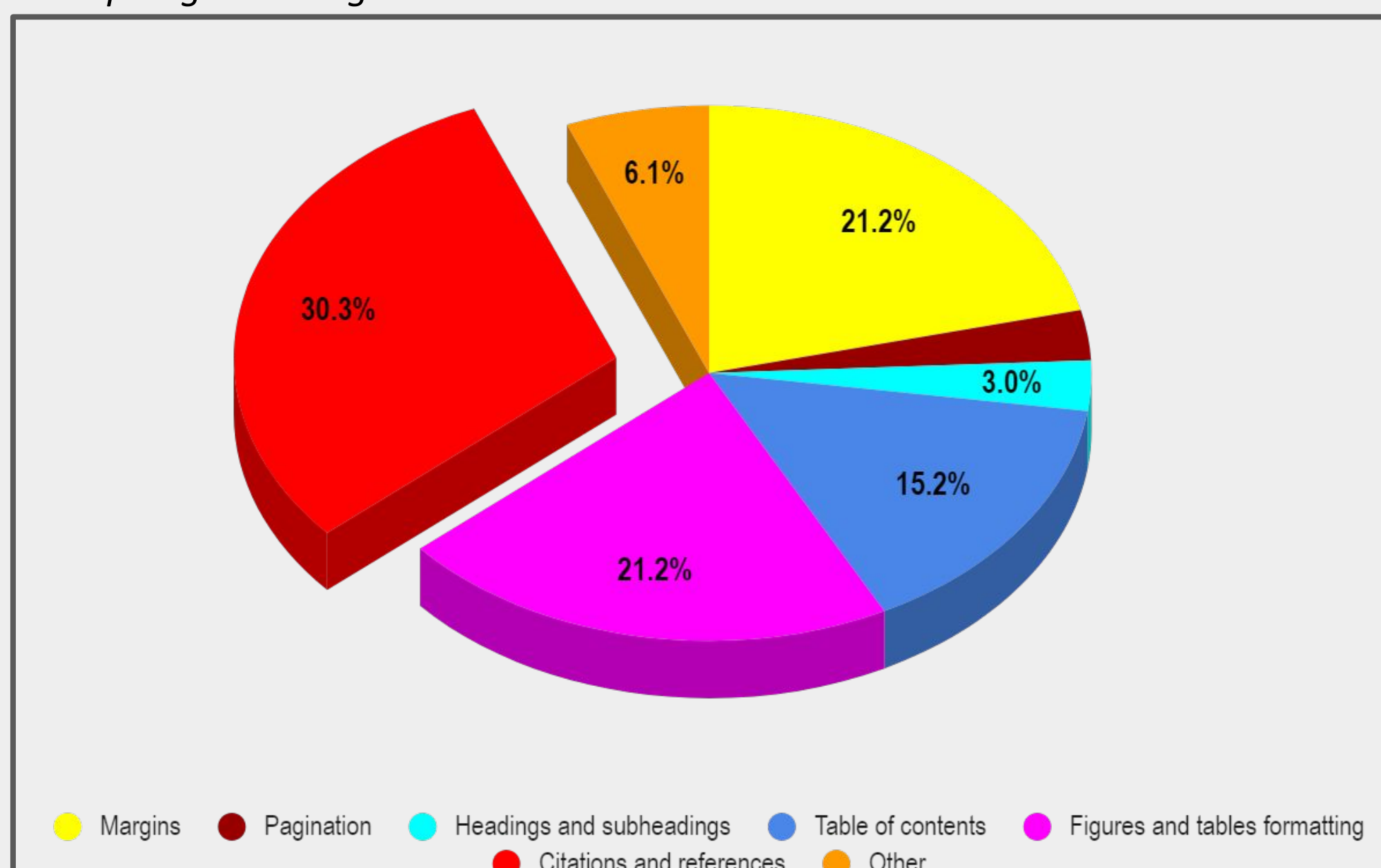
1. Below are images representing the Guidelines (Right) and Sample ETD script with parts not complying pointed out (Left).



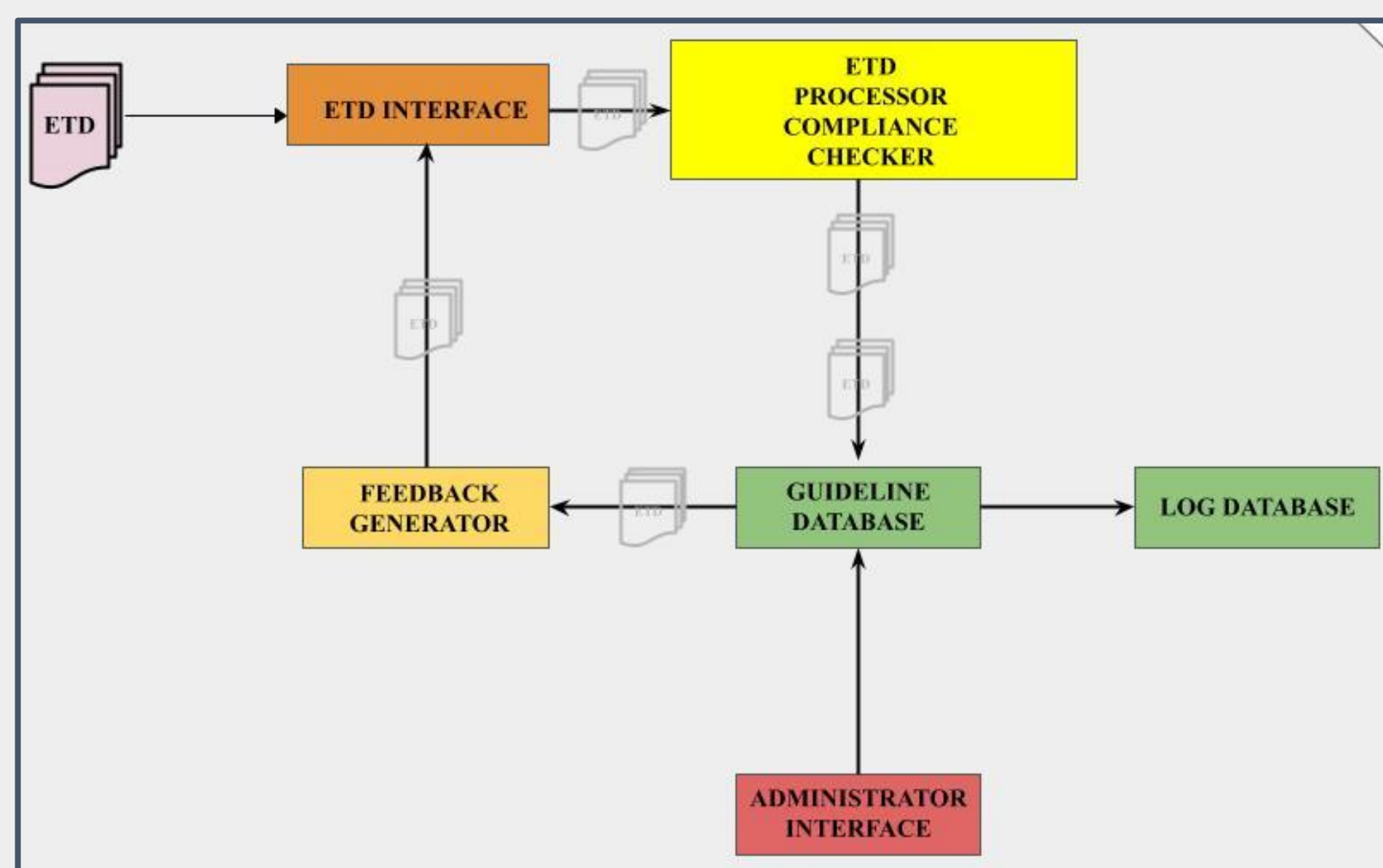
2. Below is a chart representing the average scores from the analysis of historical achieved ETD Manuscript of the entire schools at the University of Zambia on each preliminary section of the manuscript



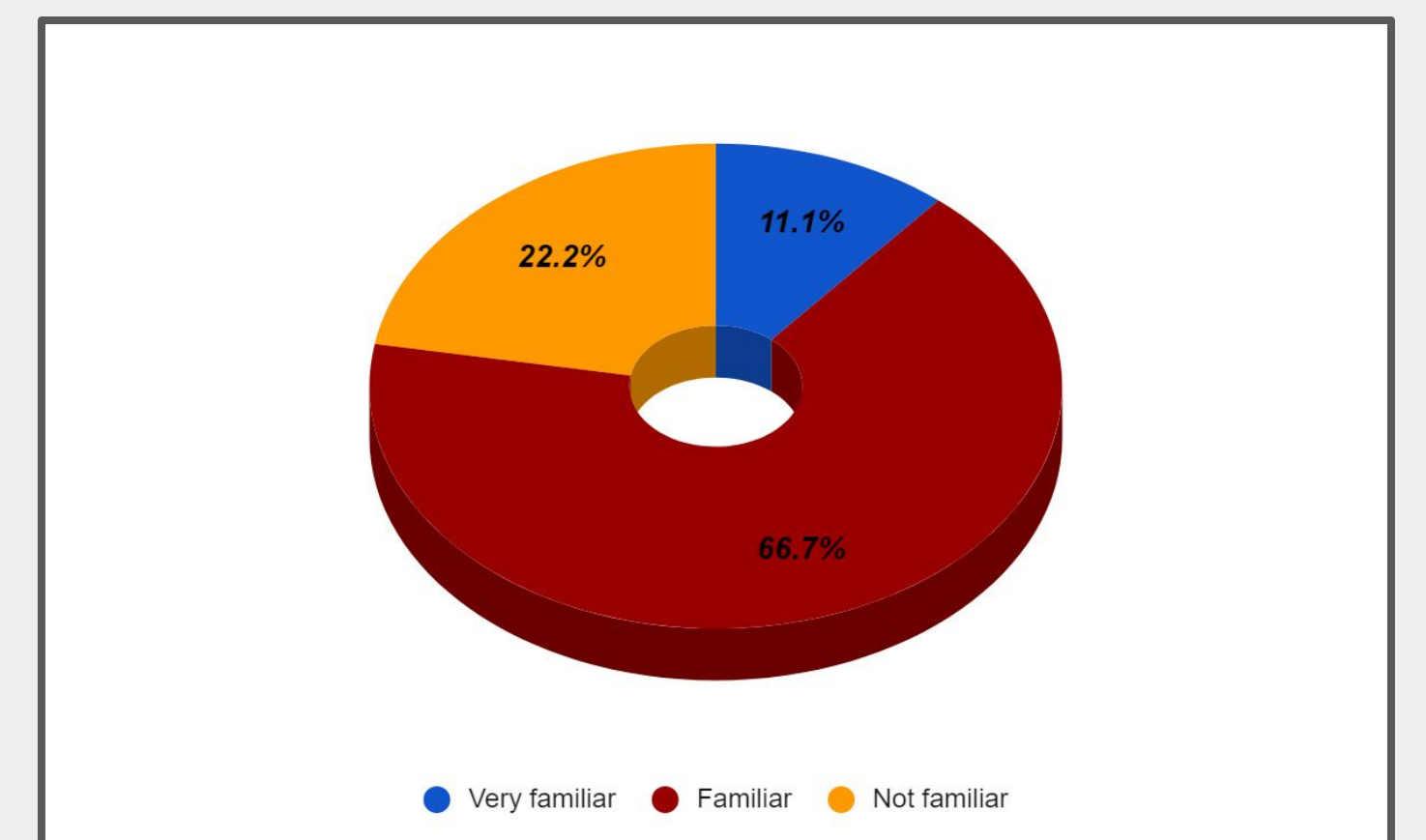
3. The chart below shows the responses gotten from Alumni Students on the challenges they faced when checking if their manuscript were in compliance with postgraduate guidelines



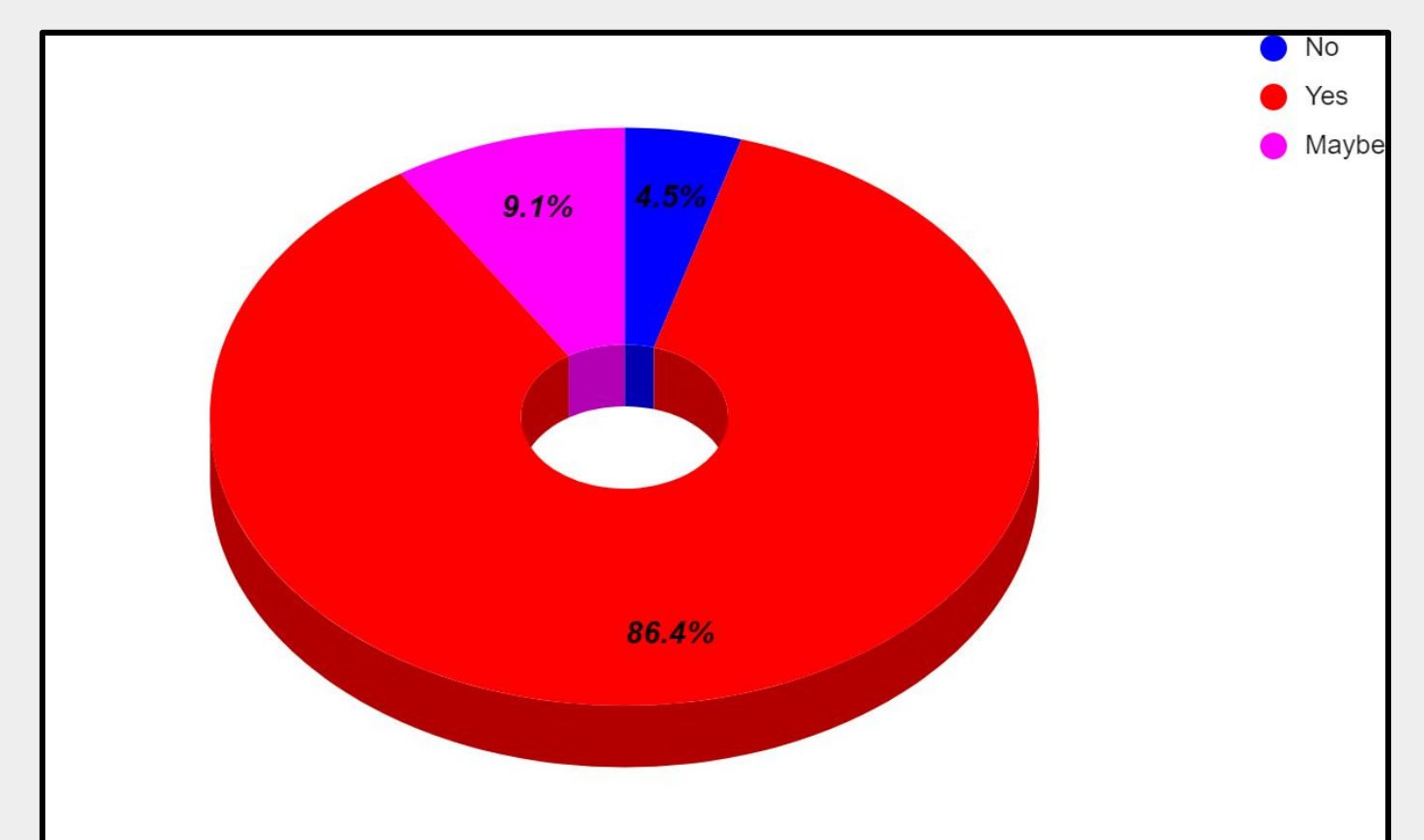
4. The diagram below illustrates how the software tool will operate once implemented.



- 3.1. The diagram on the left show responses gotten from Alumni Students when asked How familiar they were with the DRID postgraduate guidelines for ETD formatting



- 3.2. The diagram 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. 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>

