

CHAPTER I

INTRODUCTION

1.1 Background

The Faculty of Law requires an effective system to support students' academic processes, particularly internship/work practice (PKL) and undergraduate thesis title submissions. Since students must provide relevant supporting information when proposing a title, efficient communication between students and the study program is essential to ensure that the submission process complies with established regulations [1]. Currently, internship and thesis title submissions are handled manually using printed forms and handwritten approvals, resulting in lengthy processing times, limited transparency, and inefficient use of time and resources [2]. The lack of an integrated retrieval system necessitates manual verification against existing data, which makes confirming the originality of internship and thesis titles a significant difficulty. This procedure is laborious, prone to misunderstandings, and keeps students from independently confirming the names they have suggested.

SIMAS-HU, a web-based system that combines internship and thesis title submissions, was created to overcome these problems. The system facilitates more organized and successful academic administration by increasing data management effectiveness and information accessibility [3]. Students can submit titles online, monitor the status of their submissions in real time, and obtain digital approvals from approved academic personnel using SIMAS-HU. In order to provide transparent progress monitoring and more efficient oversight, the system additionally creates verified PDF documents with barcodes and keeps track of all submission modifications in an activity log.

SIMAS-HU provides an automatic title search feature that helps prevent title duplication and reduces manual verification by the KPS administrator. For quick and precise title retrieval, the search function uses the Aho-Corasick algorithm. It effectively matches title patterns, including partial word prefixes, with linear time complexity by employing a trie-based automaton [4]. Compared with other string matching algorithms, such as Boyer-Moore, which is more efficient for

long-pattern searches, *Knuth-Morris-Pratt (KMP)*, which is well suited for repetitive patterns, and *Rabin-Karp*, which relies on hashing 2³² susceptible to hash collisions, the *Aho-Corasick* algorithm is considered more appropriate for title searching in academic environments because it can process multiple patterns simultaneously while maintaining high efficiency and stability [5].

In addition to the *Aho-Corasick* algorithm, this study also employs the *BM25* method as a subsequent ranking layer to rank the search results for internship and thesis titles. *BM25* was selected because of its ability to provide stable and accurate relevance scores in information retrieval systems [6]. Compared with conventional methods such as *TF-IDF*, which assign static term weights without considering document length, *BM25* has been shown to produce more accurate search results by normalizing variations in document length [7]. In addition to keyword-based matching, this study also utilizes a word embedding method known as *Sentence-BERT (SBERT)*, which is capable of identifying semantic similarity even when students use different terms. With this approach, the system can detect semantically similar titles. For example, the title “Consumer Protection” can be recognized as semantically related to “Buyers’ Rights” even though the two titles do not share the same keywords [8].

SIMAS-HU was tested through the Black Box Testing technique, which examines the system's inputs and outputs while disregarding how the program code is internally structured, in order to confirm that the system satisfies user requirements. This method was chosen because it better suits the evaluation of the system's functional aspects from an end-user standpoint. The system was also evaluated using the *Normalized Discounted Cumulative Gain (NDCG)*, *Mean Reciprocal Rank (MRR)*, and *Recall* metrics to measure the effectiveness of the search functionality, as these metrics reflect both the relevance and completeness of the search results. Research findings indicate that a high *NDCG* score represents good ranking quality based on document relevance, the *MRR* score reflects the system’s ability to present relevant documents in the top search results, while the *Recall* score indicates the system’s effectiveness in retrieving all documents relevant to a user’s query [9]. The use of *NDCG*, *MRR*, and *Recall* as assessment metrics offers a more thorough evaluation of search quality in SIMAS-HU. By

integrating online title submission, student activity logs, title search based on the Aho-Corasick algorithm, BM25, and Sentence-BERT (SBERT), together with comprehensive system testing, SIMAS-HU is anticipated to significantly contribute to enhancing the efficiency and transparency of academic operations in the Faculty of Law.

1.2 Problem Formulation

Based on the background presented above, the main research problems addressed in this study can be formulated as follows:

1. How can a web-based information system be designed to support the internship and thesis title submission process in the Faculty of Law in order to improve the efficiency and effectiveness of academic administration?
2. How can an information retrieval system for internship and thesis data be designed and developed using the *BM25* and *Sentence-BERT (SBERT)* methods based on the *Aho-Corasick* algorithm to perform title-based searches?
3. What are the results of the system's functional testing using the *Black Box Testing* method and the performance evaluation of the internship and thesis search feature using the *NDCG*, *MRR*, and *Recall* metrics?

1.3 Research Objectives

Based on the research problems, the objectives of this study are as follows:

1. To develop a web-based information system that supports the internship and thesis title submission process in the Faculty of Law, thereby improving the efficiency and effectiveness of academic administration.
2. To investigate the implementation of the *Aho-Corasick* algorithm, *BM25*, and *Sentence-BERT (SBERT)* in the system for title-based retrieval of internship and thesis data as reference materials.
3. To evaluate the system's functionality using the *Black Box Testing* method and assess the performance of the internship and thesis search feature using the *NDCG*, *MRR*, and *Recall* metrics.

1.4 Research Benefit

Based on the research objectives, the expected contributions of this study are as follows:

1. To facilitate the internship and thesis title submission process for students in the Faculty of Law, thereby improving the effectiveness and efficiency of academic administration.
2. To provide insights into the implementation of the Aho-Corasick algorithm, BM25, and Sentence-BERT (SBERT) for internship and thesis title retrieval, enabling faster and more accurate search results.
3. To enhance knowledge and practical experience in the development of web-based academic information systems, complemented by *Black Box Testing* and search performance evaluation using the *NDCG*, *MRR*, and *Recall* metrics..
4. To serve as a reference for future research and development on the application of information technology to support transparency and accuracy in academic data management within higher education institutions..

1.5 Research Limitation

To ensure that the study remains focused and within its intended scope, the following research limitations are established:

1. The proposed system focuses on internship and thesis title submission, including title registration, data management, title retrieval, and digital approval document generation.
2. This research is implemented as a web-based application using the *Laravel* 11 framework, *MySQL* as the database management system, and *PHP* as the programming language.
3. The implementation of the *Aho-Corasick* algorithm, *BM25*, and *Sentence-BERT (SBERT)* in the search system is limited to title-based retrieval of internship and thesis titles and does not include searches within the full content of internship or thesis reports.
4. System testing is limited to functional testing using the *Black Box Testing* method and the evaluation of the internship and thesis search feature using the *NDCG*, *MRR*, and *Recall* metrics..