UNDERGRADUATE THESIS

DEVELOPMENT OF SEMANTIC SEARCH SYSTEM WITH QUERY EXPANSION FOR THESIS RETRIEVAL IN THE UPN ACADEMIC REPOSITORY

DELA PUSPITA LASMININGRUM
NPM 22081010209

THESIS ADVISORS
Eva Yulia Puspaningrum, S.Kom., M.Kom.
Budi Mukhamad Mulyo, S.Kom., M.T.

MINISTRY OF HIGHER EDUCATION, SCIENCE, AND TECHNOLOGY
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FACULTY OF COMPUTER SCIENCE
INFORMATICS STUDY PROGRAM
SURABAYA
2026

APPROVAL SHEET

DEVELOPMENT OF SEMANTIC SEARCH SYSTEM WITH QUERY EXPANSION FOR THESIS RETRIEVAL IN THE UPN ACADEMIC REPOSITORY

By :
DELA PUSPITA LASMININGRUM
NPM. 22081010209

Has been defended before, and accepted by, the Board of Assessors of the Thesis Examination of the Informatics Study Program, Faculty of Computer Science, Universitas Pembangunan Nasional Veteran Jawa Timur, on April 15, 2026.

Approved,

Eva Yulia Puspaningrum, S.Kom., M.Kom. NIP. 19890705 2021212 002	.....	**(Advisor I)**
Budi Mukhamad Mulyo, S.Kom., M.T. NIP. 19891118 202406 1 003	.....	**(Advisor II)**
Dr. Rizky Parlika, S.Kom., M.Kom. NIP. 19840518 2021211 003	.....	**(Head Assessor)**
Achmad Junaidi, S.Kom., M.Kom. NIP. 19781110 2025211 048	.....	**(Assessor I)**

Acknowledge by,
Dean of the Faculty of Computer Science

Prof. Dr. Ir. Novirina Hendrasarie, MT
NIP. 19681126 199403 2 001

APPROVAL SHEET

DEVELOPMENT OF SEMANTIC SEARCH SYSTEM WITH QUERY EXPANSION FOR THESIS RETRIEVAL IN THE UPN ACADEMIC REPOSITORY

By:
DELA PUSPITA LASMININGRUM
NPM. 22081010209

Approved to proceed to the Thesis Examination

Approved by,

Coordinator of Informatics Study Program
Faculty of Computer Science

Dr. Intan Yuniar Purbasari, S.Kom. MSc.
NIP. 19800602 202521 2 029

STATEMENT OF ORIGINALITY

I am the undersigned:

Student Name	: DELA PUSPITA LASMININGRUM
NPM	: 22081010209
Degree Program	: Bachelor (S1)
Study Program	: Informatics
Faculty	: Faculty of Computer Science

Hereby declares that this undergraduate thesis contains no part of any other scientific work that has been submitted to obtain an academic degree at any higher education institution. Furthermore, it does -not contain any work or opinions previously written or published by others, except for those which are explicitly cited in this thesis and listed completely in references.

And I declare that this scientific document is free from elements of plagiarism. If in the future indications of plagiarism are found in this Thesis, I am willing to accept sanctions in accordance with the applicable laws and regulations.

Thus, I made this statement without any coercion from anyone and to be used as it should.

Surabaya, May 18th, 2026

Declarant,

DELA PUSPITA LASMININGRUM

NPM. 22081010209

ABSTRACT

Student Name / NPM	:	DELA PUSPITA LASMININGRUM / 22081010209
Thesis Title	:	DEVELOPMENT OF SEMANTIC SEARCH SYSTEM WITH QUERY EXPANSION FOR THESIS RETRIEVAL IN THE UPN ACADEMIC REPOSITORY
Supervisor	:	1. Eva Yulia Puspaningrum, S.Kom., M.Kom. 2. Budi Mukhamad Mulyo, S.Kom., M.T.

Information retrieval systems play an important role in helping users efficiently find relevant documents. In higher education, thesis retrieval has become an essential need for students as a writing reference and to avoid similarity in research topics. The digital repository of UPN "Veteran" Jawa Timur provides access to student theses; however, the existing search system still relies on explicit keyword matching, making it less effective when users submit short, general queries or use different terms with similar meanings.

This study aims to design and implement a thesis retrieval system based on semantic search capable of understanding the contextual meaning of queries and documents. The proposed approach utilizes IndoSBERT to generate semantic representations (embeddings) of thesis titles, abstracts, and user queries. To improve sensitivity to term variations and limitations of short queries, an embedding-based query expansion technique is applied. The embedding retrieval process is conducted using FAISS with cosine similarity to maintain search efficiency on a large-scale dataset. The study includes system design, embedding generation, query expansion integration, as well as implementation and functional testing of the retrieval system.

The experimental results show that the system is capable of retrieving relevant documents based on semantic meaning despite vocabulary differences between queries and documents. The semantic fine-tuning model with query expansion achieved an nDCG@15 score of 0.8470, indicating that relevant documents tend to appear at the top of the search results. The application of fine-tuning and query expansion proved effective in improving the quality of semantic representations and enriching query context, enabling the system to capture semantic relationships better than the baseline semantic approach. In terms of performance, the system was able to process searches over 15,326 documents with an average retrieval time of approximately 0.5 seconds using FAISS. Therefore, the IndoSBERT-based semantic search approach combined with fine-tuning and query expansion can serve as an alternative approach for thesis retrieval systems, particularly for queries with vocabulary variations.

Keywords: *semantic search*, IndoSBERT, *query expansion*, FAISS, thesis retrieval.

ACKNOWLEDGEMENTS

Praise and gratitude are rendered unto Allah SWT for His abundant grace and guidance, which have enabled the author to complete this thesis entitled "Development of Semantic Search System with Query Expansion for Thesis Retrieval in the UPN Academic Repository". This thesis is prepared as one of the academic requirements for obtaining the degree of Bachelor of Computer Science in the Informatics Study Program, Faculty of Computer Science, Universitas Pembangunan Nasional "Veteran" Jawa Timur.

The process of preparing this research was not without its challenges; however, through the guidance, support, and prayers of various parties, all obstacles were overcome. Therefore, with sincere humility, the author wishes to express deep appreciation and heartfelt gratitude to:

1. Prof. Dr. Ir. Novirina Hendrasarie, MT, as Dean of the Faculty of Computer Science, Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Dr. Intan Yuniar Purbasari, S.Kom., MSc., as Head of the Informatics Study Program, Faculty of Computer Science Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Eva Yulia Puspaningrum, S.Kom., M.Kom. and Budi Mukhamad Mulyo, S.Kom., M.T., as academic supervisors who generously dedicated their time to providing guidance, serving as critical brainstorming partners, and assisting the author in identifying effective solutions to every technical challenge encountered throughout the thesis process.
4. Dr. Rizky Parlika, S.Kom., M.Kom. and Achmad Junaidi, S.Kom., M.Kom., as thesis examiners who provided invaluable feedback in evaluating and improving the various shortcomings of this research in pursuit of a higher quality thesis.
5. Retno Mumpuni, S.Kom., M.Sc., as academic advisor who has provided academic direction and guidance to the author from the beginning of the study period through to the final stage of completion.
6. All Lecturers of the Informatics Study Program, who have equipped the author with knowledge, insight, and inspiration throughout the course of study.

7. The Coordinator and staff of the UPT Library of UPN "Veteran" Jawa Timur, who granted research permission and facilitated the author in the data collection process essential to this research.
8. All my respondents from various majors in UPN Jatim, who voluntarily participated in the testing process. Thank you for generously dedicating your time and providing the valuable feedback necessary for the evaluation of this research.
9. My beloved Mother and Father, who have ceaselessly provided material and spiritual support, along with my Sister as my greatest inspiration. My deepest gratitude is extended for all the prayers, affection, and unwavering material and spiritual support that sustained the author through to the successful completion of this thesis.
10. Kevin Ricky Pradana, my greatest support system who has consistently stood by the author throughout the challenges of completing this thesis. Thank you for being an exceptional discussion partner, for accompanying exploring various coffeeshops to meet our daily targets, and for the endless mutual motivation as we strive to graduate and achieve our dreams together.
11. Izza, Firli, Dilla, Aileen, Rahma, and Dwi, as the author's closest friends and fellow thesis fighters. Thank you for sharing the burdens and joys of this journey together.
12. My Zfordda Mojokerto friend circle, who, despite our occasional gatherings, have always provided deep conversations and valuable insights that helped the author gain new perspectives on this thesis.
13. Members of HIMATIFA, particularly the Executive Board (BPH) and Heads of Departments for the 2023 and 2024 periods. Thank you for the extensive insights and knowledge shared across all matters, which significantly contributed to the author's personal growth and the completion of this research.
14. All other friends and parties who cannot be mentioned individually, but have provided support and assistance in any form during the author's academic journey.

The author acknowledges that this thesis contains many shortcomings. Therefore, constructive criticism and suggestions from all parties are greatly welcomed in the interest of improving this work. Finally, with all the limitations the author possesses, it is hoped that this report may be of benefit to all parties in general, and to the author in particular.

Surabaya, May, 18th 2026

Penulis

TABLE OF CONTENTS

LIST OF FIGURE

LIST OF TABLES

LIST OF FORMULA