

INTISARI

Penelitian ini bertujuan untuk merancang dan mengimplementasikan FAQ *Retrieval Chatbot* Penerimaan Mahasiswa Baru (PMB) pada *website* STMIK AMIKOM Surakarta menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF) dan *Cosine similarity*. Sistem dikembangkan untuk membantu penyampaian informasi PMB secara otomatis dengan mengambil jawaban dari basis pengetahuan FAQ yang telah disusun dan divalidasi. Basis pengetahuan final terdiri atas 1.010 FAQ dalam enam kategori, yang menghasilkan 5.587 dokumen *corpus* dan 12.785 *vocabulary* TF-IDF. Sistem diimplementasikan menggunakan *WordPress* sebagai antarmuka pengguna, *FastAPI* sebagai *backend* REST API, *MySQL* sebagai basis data, dan *Virtual Private Server* (VPS) sebagai lingkungan implementasi *backend*. Mekanisme pemrosesan dilakukan secara bertahap melalui *Rule-based*, TF-IDF dan *Cosine similarity*, *Fuzzy Matching*, serta *Escalation*. *PreProcessing* teks meliputi *case folding*, pembersihan karakter *non-alfanumerik*, normalisasi kata informal, dan normalisasi spasi. Sistem juga menyediakan *dashboard* admin untuk pengelolaan FAQ, *reIndex*, percakapan, *lead*, *monitoring*, dan analitik. Hasil *regression test* terhadap 30 kasus uji memperoleh 30 *content PASS* dan seluruh *routing Layer* sesuai. *Confusion Matrix* terhadap 50 pertanyaan menghasilkan *Accuracy* 82,00%, *Precision* 76,92%, *Recall* 100,00%, dan *F1-Score* 86,96%. Pengujian *5-fold Cross Validation* terhadap 4.170 variasi pertanyaan menghasilkan *Exact top-1 Accuracy* $33,02\% \pm 1,02\%$, sedangkan analisis *Answer-equivalent Top-5* menghasilkan $86,59\% \pm 1,02\%$. *User Acceptance Test* menghasilkan 100% skenario *fungsi*ional diterima, dan *USE Questionnaire* memperoleh nilai 87,50% dengan kategori Sangat Layak. Hasil penelitian menunjukkan bahwa sistem dapat digunakan untuk mendukung layanan informasi PMB, meskipun kemampuan generalisasi terhadap variasi pertanyaan baru dan pengelolaan FAQ yang memiliki informasi serupa masih perlu ditingkatkan.

Kata kunci: TF-IDF, *Cosine similarity*, FAQ *Retrieval*, *Chatbot*, Penerimaan Mahasiswa Baru.

ABSTRACT

This study aims to design and implement a New Student Admission (PMB) FAQ Retrieval Chatbot on the STMIK AMIKOM Surakarta website using Term Frequency-Inverse Document Frequency (TF-IDF) and Cosine similarity. The system was developed to support automated PMB information services by retrieving answers from a validated FAQ knowledge base. The final knowledge base consists of 1,010 FAQs in six categories, producing 5,587 corpus documents and 12,785 TF-IDF vocabulary terms. The system was implemented using WordPress as the user interface, FastAPI as the REST API backend, MySQL as the database, and a Virtual Private Server (VPS) as the backend deployment environment. The chatbot applies a staged Processing mechanism consisting of Rule-based Processing, TF-IDF and Cosine similarity, Fuzzy Matching, and Escalation. Text PreProcessing includes case folding, removal of non-alphanumeric characters, informal-word normalization, and whitespace normalization. An administrative dashboard is also provided for FAQ management, reIndexing, conversation history, lead management, monitoring, and analytics. The regression test involving 30 critical test cases resulted in 30 content PASS results with all routing Layers functioning as expected. The Confusion Matrix evaluation on 50 questions achieved an Accuracy of 82.00%, Precision of 76.92%, Recall of 100.00%, and F1-Score of 86.96%. The 5-fold Cross Validation on 4,170 question variations produced an Exact top-1 Accuracy of $33.02\% \pm 1.02\%$, while the Answer-equivalent Top-5 analysis achieved $86.59\% \pm 1.02\%$. The User Acceptance Test resulted in 100% functional scenario acceptance, while the USE Questionnaire achieved 87.50%, categorized as Very Feasible. The results indicate that the system can support PMB information services, although its generalization capability for unseen question variations and the management of semantically overlapping FAQs still require further improvement.

Keyword: TF-IDF, Cosine similarity, FAQ Retrieval, Chatbot, New Student Admission.