

ABSTRACT

The demand for localized information in Lhokseumawe remains sub-optimal due to the generalized nature of current news search engines, which lack specific spatial filtering mechanisms at the city level. To address this issue, this study aims to develop an Artificial Intelligence (AI)-driven local news retrieval and summarization system. The proposed pipeline seamlessly integrates web scraping, Named Entity Recognition (NER), extractive summarization, and semantic similarity matching. Built using Python and the Flask framework, the system utilizes HTTPX for primary data extraction and Selenium as a fallback mechanism. The collected articles are filtered using a spaCy-based NER model to isolate the "Lhokseumawe" location entity. Subsequently, the filtered content is condensed using a Sentence Transformer model and ranked according to its semantic relevance to the user's input. Evaluation results demonstrate the efficacy of the NER model in spatial filtering, achieving a Precision of 80.49%, a Recall of 91.67%, and an F1-Score of 85.72%. Furthermore, qualitative assessments indicate that the generated summaries successfully deliver concise and highly relevant core information. The system dynamically adapts its output format, providing narrative texts for general information searches and structured lists for recommendation-based queries. Ultimately, this research proves that combining NER, extractive summarization, and semantic matching provides an effective framework for presenting location-specific news content.

Keywords: *Artificial Intelligence (AI), extractive summarization, Named Entity Recognition (NER), semantic similarity, web scraping*