

Peningkatan Akurasi Transliterasi Reversibel Pegon - Latin dengan Pendekatan Rule-Based dan Model Naive Bayes dengan Metode Tokenisasi Bigram = Improving the Accuracy of Reversible Transliteration from Pegon to Latin using a Rule-Based Approach and Naive Bayes Model with Bigram Tokenization Method

Adrian Larry Ananda Sudarthio, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920564099&lokasi=lokal>

Abstrak

Indonesia memiliki kekayaan warisan budaya yang berlimpah, salah satunya adalah aksara Pegon, sebuah modifikasi aksara Arab yang tumbuh dan berkembang pada zaman penyebaran agama Islam di Nusantara. Kekayaan tersebut tentu tak dapat dilepaskan pada keaktifan menulis para ulama sehingga terbentuklah kekayaan manuskrip Pegon yang kaya. Namun dikarenakan masyarakat Indonesia lebih paham aksara Latin, tidak banyak yang bisa membaca aksara Pegon, bahkan yang memahami aksara Arab pun belum tentu dapat mengerti Pegon. Oleh karena itu, dibutuhkan transliterasi antara aksara Pegon dan Latin. Penelitian ini berfokus pada peningkatan akurasi model transliterasi rule-based yang telah dikembangkan sebelumnya pada mata kuliah Proyek Perangkat Lunak (PPL) genap 2022/2023. Lalu, penelitian ini juga berfokus pada pengembangan transliterasi dengan model Naive Bayes berbasis tokenisasi bigram yang mengacu pada pedoman transliterasi Arab-Latin dari Kementerian Agama tahun 1987, Hasil Kongres Aksara Pegon 2022, dan SNI 9047:2023. Hasil uji coba model rule-based yang telah direvisi dengan perbaikan 15 rules dari total 87 rules menghasilkan akurasi 98.8%. Sedangkan model bigram yang telah dikembangkan menghasilkan akurasi 94%. Dua model tersebut terbukti memiliki hasil yang jauh lebih baik dibandingkan dengan model rule-based sebelumnya yang hanya memiliki akurasi 60% saja.

.....Indonesia is rich in cultural heritage, one of which is the Pegon script, a modification of the Arabic script that grew and developed during the spread of Islam in the archipelago. This wealth is undoubtedly linked to the writing activities of religious scholars, leading to the formation of a wealth of Pegon manuscripts. However, due to the Indonesian public's better understanding of the Latin script, not many can read the Pegon script, and even those who understand the Arabic script may not necessarily understand Pegon. Therefore, a transliteration between the Pegon and Latin scripts is needed. This study focuses on improving the accuracy of a previously developed rule-based transliteration model in the Software Project (PPL) course in the even semester of 2022/2023. Then, this study also focuses on the development of transliteration with a Naive Bayes model based on bigram tokenization, referring to the Arabic-Latin transliteration guidelines from the Ministry of Religion in 1987, Kongres Aksara Pegon 2022 results, and SNI 9047:2023. The revised rule-based model, with corrections to 15 rules out of a total of 87, achieved an accuracy of 98.8%. Meanwhile, the developed bigram model achieved an accuracy of 94%. These two models have proven to yield much better results than the previous rule-based model, which only had an accuracy of 60%.