

Transliterasi Pegon-Latin dengan pendekatan Data-Driven menggunakan metode Sequence-to-Sequence = Pegon-Latin Transliteration with A Data-Driven approach using Sequence-to-Sequence methods

Fitri 'Aliyah, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920553287&lokasi=lokal>

Abstrak

Indonesia kaya akan warisan budaya, salah satunya adalah aksara Pegon. Aksara ini merupakan sistem penulisan yang berkembang selama penyebaran Islam di nusantara. Warisan budaya ini adalah adaptasi dari aksara Arab yang sering digunakan oleh para ulama Islam dalam penulisan manuskrip di masa lalu. Namun, seiring dominasi aksara Latin di Indonesia, penggunaan aksara Pegon saat ini semakin berkurang, hanya pada kalangan tertentu, sehingga menyebabkan sedikitnya individu yang mampu membaca aksara Pegon. Oleh karena itu, transliterasi antara aksara Pegon ke Latin diperlukan. Penelitian ini berfokus pada pengembangan transliterasi menggunakan pendekatan berbasis data dengan model sequence-to-sequence, mengikuti pedoman transliterasi Arab-Latin dari Kementerian Agama tahun 1987, hasil Kongres Aksara Pegon 2022, dan SNI 9047:2023. Hasil penelitian menunjukkan bahwa pendekatan berbasis data menggunakan metode sequence-to-sequence mencapai akurasi 99.76% dan CER 0.010624 untuk dataset bahasa Sunda dengan model terbaik BiGRU-Att, akurasi 99.31% dan CER 0.029255 untuk dataset bahasa Jawa dengan model terbaik BiGRU-Att, dan akurasi 99.58% dan CER 0.020466 untuk dataset gabungan bahasa dengan model BiLSTM-Att. Dari hasil ini, dapat dikatakan bahwa hasil prediksi tergolong baik dengan nilai akurasi di atas 70%, nilai loss mendekati 0, dan nilai Character Error Rate (CER) mendekati 0 untuk semua dataset.

.....Indonesia, rich in cultural heritage, includes Pegon script, a writing system that evolved during the spread of Islam in the archipelago. This cultural heritage is an adaptation of the Arabic script often used by Islamic scholars in manuscript writing in the past. However, the current use of Pegon script is less popular compared to the past due to the dominance of the Latin script in Indonesia, resulting in few individuals being able to read Pegon script. Therefore, transliteration between Pegon and Latin scripts is necessary. The research concentrates on developing transliteration using a data-driven approach with sequence-to-sequence models, following the Arabic-Latin transliteration guidelines from the Ministry of Religious Affairs in 1987, the results of the Pegon Script Congress 2022, and SNI 9047:2023. The results show that the data-driven approach using the sequence- to-sequence method achieves an accuracy of 99.76% and a CER of 0.010624 for the Sundanese dataset with the best model BiGRU-Att, an accuracy of 99.31% and a CER of 0.029255 for the Javanese dataset with the best model BiGRU-Att, and an accuracy of 99.58% and a CER of 0.020466 for the combined language dataset with the BiLSTM-Att model. From these results, it can be said that the prediction results are classified as good with accuracy values above 70%, loss values close to 0, and Character Error Rate (CER) values close to 0 for all datasets.