

SISTEM EKTRAKSI INFORMASI LIRIK LAGU DAERAH MENGGUNAKAN *TERM FREQUENCY– INVERSE DOCUMENT FREQUENCY* DAN *COSINE SIMILARITY*

Nama Mahasiswa : Nur Ifdayanti
NIM : 6304221418
Dosen Pembimbing : Fajar Ratnawati, M.Kom

ABSTRAK

Lagu daerah merupakan salah satu warisan budaya Indonesia yang mengandung nilai sejarah dan kearifan lokal, namun masih banyak lirik lagu daerah yang sulit diakses dan dipahami secara luas. Penelitian ini bertujuan membangun sistem ekstraksi informasi lirik lagu daerah menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF) dan *Cosine Similarity* untuk mendukung pencarian lirik berdasarkan kata kunci. Sistem dikembangkan melalui tahapan *preprocessing* teks, pembobotan kata dengan TF-IDF, serta pengukuran kemiripan query dan dokumen menggunakan *Cosine Similarity*. Selain itu, sistem dilengkapi analisis linguistik menggunakan *POS Tagging* dan *rule-based extraction* untuk mengidentifikasi elemen seperti nomina, verba, subjek, objek, negasi, dan emosi. Hasil pengujian menunjukkan sistem memperoleh Top-1 Accuracy sebesar 100% pada 10 query uji. Evaluasi ekstraksi linguistik menghasilkan precision 0,74, recall 0,76, dan F1-score 0,75. Sistem ini mampu mengekstraksi informasi lirik lagu daerah secara efektif sebagai sarana pelestarian budaya dan edukasi digital.

Kata Kunci: *TF-IDF, Cosine Similarity, Ekstraksi Informasi, Lirik Lagu Daerah, POS Tagging.*

INFORMATION EXTRACTION SYSTEM REGIONAL SONG LYRICS USING TERM FREQUENCY – INVERSE DOCUMENT FREQUENCY AND COSINE SIMILARITY

Student name : Nur Ifdayanti
NIM : 6304221418
Advisor : Fajar Ratnawati, M.Kom

ABSTRACT

Traditional regional songs are one of Indonesia's cultural heritages that contain historical values and local wisdom. However, many regional song lyrics are still difficult to access and widely understand. This study aims to develop an information extraction system for regional song lyrics using the Term Frequency–Inverse Document Frequency (TF-IDF) and Cosine Similarity methods to support keyword-based lyric searching. The system was developed through text preprocessing stages, term weighting using TF-IDF, and similarity measurement between user queries and lyric documents using Cosine Similarity. In addition, the system is equipped with linguistic analysis using POS Tagging and rule-based extraction to identify elements such as nouns, verbs, subjects, objects, negation, and emotions. The test results show that the system achieved a Top-1 Accuracy of 100% on 10 test queries. The linguistic extraction evaluation produced a precision of 0.74, recall of 0.76, and an F1-score of 0.75. Therefore, this system is able to effectively extract information from regional song lyrics as a tool for cultural preservation and digital education.

Keywords: *TF-IDF, Cosine Similarity, Information Extraction, Traditional Regional Song Lyrics, Part-of-Speech Tagging.*