



SISTEM TEMU KEMBALI INFORMASI
KELOMPOK 10

ANALISIS DAN CLUSTERING MAKNA AYAT AL-QUR'AN DENGAN MINI SEARCH ENGINE BM-25

➤ **MUHAMMAD RIZKY RAMADHAN**
➤ **DHEAZ KELVIN HARAHAAP**
➤ **DAFFA PRATAMA YUDHISTIRA**

➤ **REYNALDO ARYA BUDI TRISNA**
➤ **MUHAMMAD AZKA PINANDITO**

PENDAHULUAN

Penelitian ini bertujuan untuk mengembangkan program komputasi yang memungkinkan pencarian dan analisis ayat Al-Quran berdasarkan tema dan topik. Metode yang digunakan meliputi Vector Space Model, K-Means Clustering, dan Latent Dirichlet Allocation (LDA). Program ini memungkinkan pengguna untuk mencari ayat-Al-Quran yang relevan dengan pertanyaan atau topik yang spesifik. Penelitian ini akan menyajikan hasil dan analisis dari pencarian ayat-Al-Quran menggunakan metode-metode tersebut. Tujuan penelitian ini adalah untuk memberikan wawasan baru dalam pemahaman dan analisis komputasional Al-Quran.

ALGORITHM



III.1. Preprocessing Teks

Preprocessing merupakan tahap penting sebelum mengelompokkan dokumen. Proses ini bertujuan untuk memudahkan pengolahan dan meningkatkan akurasi. Tahapan preprocessing yang digunakan meliputi:

1. Case folding: Mengubah semua huruf dalam dokumen menjadi huruf kecil agar kata-kata dianggap sama dalam perhitungan.
2. Cleaning: Menghapus karakter yang dianggap tidak relevan, seperti tanda baca dan angka.
3. Tokenization: Memotong kalimat menjadi kata-kata dalam setiap dokumen.
4. Filtering: Memilih kata-kata penting dari hasil tokenisasi.
5. Stop list: Menghilangkan kata-kata yang tidak penting.
6. Stemming: Mengubah kata-kata dengan imbuhan tambahan menjadi kata dasar.

III.2. Term Frequency-Inverse Document Frequency (TF-IDF)

TF-IDF adalah algoritma yang menghitung frekuensi suatu term dalam dokumen dan inverse frekuensi dokumen yang mengandung term tersebut. Setelah menghitung TF-IDF, perlu dilakukan normalisasi agar nilai tetap antara 0 dan 1.

III.3. Clustering: K-Means

Clustering merupakan metode penting dalam pembelajaran mesin yang digunakan di berbagai bidang. Algoritma clustering memberikan cara cepat untuk memahami data. Salah satu metode clustering yang populer adalah K-Means. Tujuannya adalah mencapai kesamaan yang tinggi di dalam klaster dan kesamaan yang rendah antar klaster. Dalam K-Means, objek yang saling dekat dianggap sebagai klaster. Algoritma ini memiliki kelebihan dalam kesederhanaan, efisiensi, dan kecepatan. Namun, algoritma ini sensitif terhadap titik awal dan pemilihan sampel awal yang berbeda dapat menghasilkan hasil yang berbeda. Selain itu, algoritma ini menggunakan metode gradien dalam pencarian titik pusat kelompok, yang dapat menyebabkan terjebak pada titik minimum lokal jika titik awal tidak tepat.

METODE

III.4. Metode Elbow

Metode Elbow adalah metode visual untuk menentukan jumlah cluster terbaik dengan membandingkan perbedaan Sum of Square Error (SSE) dari setiap cluster. Pada grafik SSE, perbedaan yang paling signifikan akan membentuk sudut siku yang menunjukkan jumlah cluster terbaik. Metode ini digunakan untuk optimasi penentuan jumlah cluster terbaik, meskipun pemilihan awal pusat cluster masih acak. Namun, penentuan jumlah cluster terbaik dengan metode Elbow dapat membutuhkan waktu lebih lama karena jumlah iterasi yang lebih banyak untuk mencapai kesamaan pola yang terbentuk.

III.5. Jarak Centroid

Jarak antara suatu instance dan centroid klaster dari setiap kelas dapat menggambarkan afinitas antara data klaster, yang berguna dalam konstruksi model dalam fase kolaborasi. Jarak centroid dapat diukur menggunakan beberapa metode, seperti jarak Euclidean, jarak Mahalanobis, atau jarak geodesik dari manifold.

III.6. Metode Pencarian

BM-25 Metode pencarian atau searching method adalah strategi yang digunakan untuk mencari informasi atau data spesifik dalam suatu konteks. Dalam komputasi, metode pencarian sering digunakan dalam pemrograman atau sistem informasi untuk menemukan data dalam struktur yang lebih besar, seperti array, daftar, atau basis data.

BM-25 (Okapi Best Matching 25) adalah fungsi peringkat berbasis bag-of-words yang digunakan untuk mengestimasi relevansi suatu dokumen. BM-25 juga merupakan model pembelajaran mesin, termasuk model neural yang kuat. BM-25 dianggap sebagai formula terbaik dalam kelas best match karena efektivitasnya dalam mengurutkan dokumen berdasarkan query yang dicari.

KONTRIBUSI

KONTRIBUSI JURNAL DI ATAS ADALAH MEMBAHAS TENTANG PENGEMBANGAN SISTEM Pencarian dan Pengelompokan Terjemahan Al-Quran. Jurnal ini memberikan informasi tentang tahap-tahap dalam penelitian, seperti teknik pengambilan data, metodologi penelitian, dan tahap preprocessing data.





SISTEM TEMU KEMBALI INFORMASI
KELOMPOK 10

HASIL DAN PEMBAHASAN



<https://colab.research.google.com/drive/1w6txAWSA23d2oFMnNhUprUp1PdZrP0u-?usp=sharing>





KELOMPOK 10

KESIMPULAN



SISTEM TEMU KEMBALI INFORMASI

THANK YOU



KELOMPOK 10