Nama:

* Ferdinandus Lembabang Sula (165314046)
* Christianto Ega Prasetyo (165314051)

Template Laporan Aplikasi Pengelompokan Dokumen

Judul: Pengelompokkan document lirik lagu menggunakan K-means (Document dikelompokkan berdasarkan 2 class document lagu, yaitu lagu Rohani dan lagu Pop)

Implementasi Algoritma K-Means untuk mengelompokkan Dokumen berbasis TF-IDF dan Cosine Similarity

Kasus : Pengelompokkan dokumen lirik lagu dengan menggunakan algoritma k-means

1. Pendahuluan

* Apa itu pengelompokan dokumen?

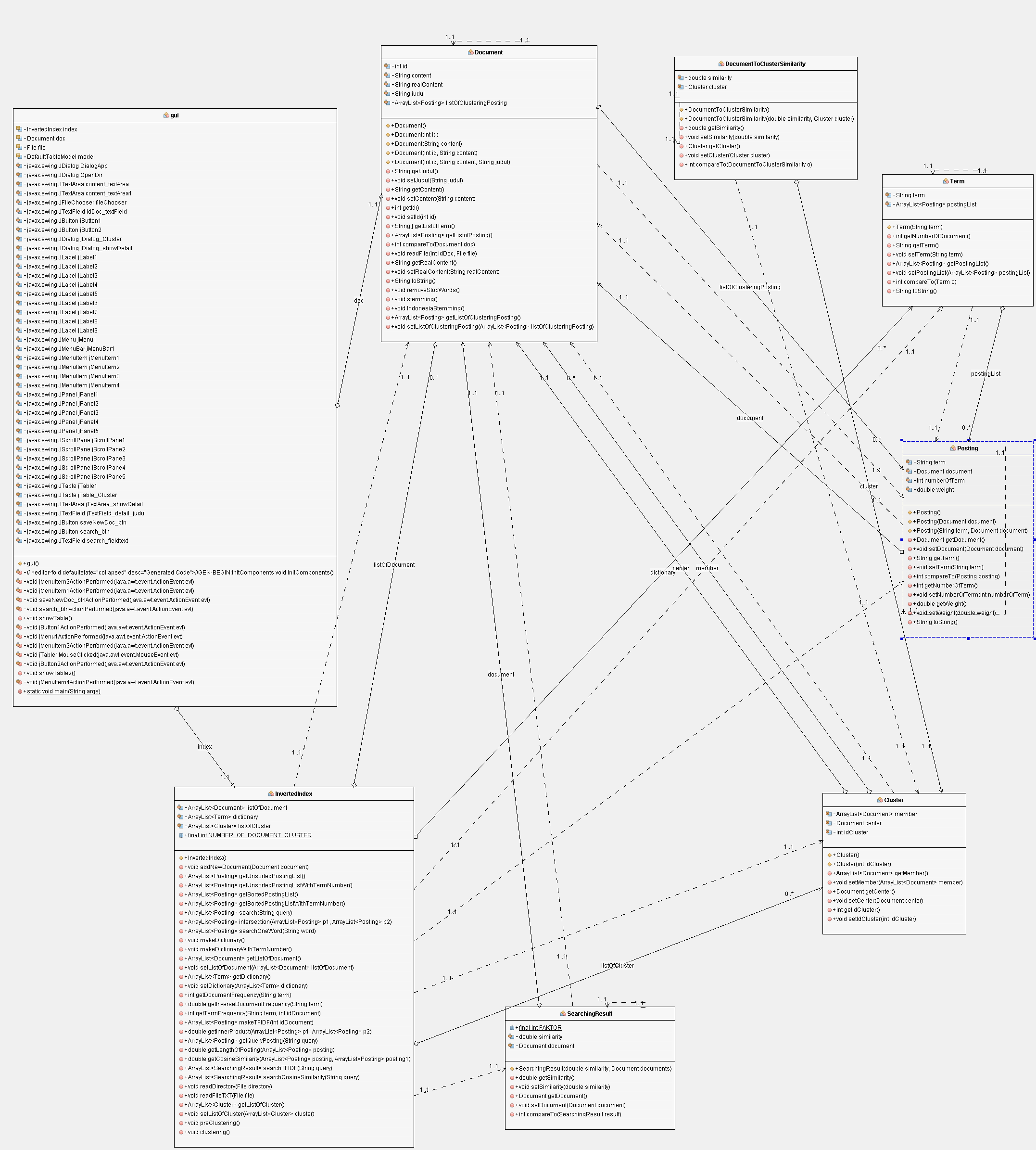
Dokumen clustering (pengelompokan dokumen atau Text) adalah dokumen organisasi otomatis, ekstraksi topik dan pengambilan informasi yang cepat atau penyaringan.

Metode Clustering yang dapat digunakan untuk secara otomatis mengelompokkan dokumen diambil ke dalam daftar kategori bermakna, untuk mengelompokkan dokumen tersebut kami menggunakan algoritma K-MEANS.

* TF-IDF adalah suatu algoritma atau formula yang digunakan untuk menghitung bobot masing-masing dokumen terhadap kata kunci. Term Frekuensi \* Inverse Document Frequency atau biasa disingkat dengan tf-idf. Term Frequency (TF) adalah frekuensi dari kemunculan sebuah term dalam dokumen yang bersangkutan. Inverse Document Frequency (IDF) adalah suatu statistic yang mengkarakteristikan sebuah term dalam keseluruhan dokumen
* Cosine Similarity adalah metode untuk menghitung kesamaan antara dua buah objek yang dinyatakan dalam dua buah vector dengan menggunakan kata kunci dari sebuah dokumen sebagai ukuran

1. Alat dan Bahan

* Bahan teks yang digunakan yaitu lirik lagu berupa file txt, ada 2 genre lagu (lagu rohani 24 file dan pop 50 file total file txt = 74) yang mana lirik diambil dari website-website yang menyediakan lirik-lirik lagu
* Aplikasi berbasis Java Dekstop



* Precision adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh system. Sedangkan Recall adalah tingkat keberhasilan system dalam menemukan kembali sebuah informasi

1. Evaluasi dan Uji coba
2. Bagi dokumen anda dalam 2 kategori besar. Asumsi dokumen akan dikelompokkan dalam 2 kluster.
3. Lakukan pengelompokan menggunakan aplikasi. Hasil pengelompokan anda catat dan rekam. Lalu bandingkan hasil dengan langkah a.

Masukan system:

Pusat kluster

|  |  |  |  |
| --- | --- | --- | --- |
| Dokumen | Actual | Prediksi | Percobaan 1  Error |
| Aku Percaya (2) | 0 | 0 | 0 |
| Allah Peduli (4) | 0 | 0 | 0 |
| Bapa yang Kekal (9) | 0 | 0 | 0 |
| Begawai Rajawali (10) | 0 | 0 | 0 |
| Belum Ada Judul(11) | 1 | 0 | 1 |
| Betapa hatiku(14) | 0 | 0 | 0 |
| Dalam lembah kelam(21) | 0 | 0 | 0 |
| Dengan apa kan ku Balas(23) | 0 | 0 | 0 |
| Doa Pengobral Dosa(24) | 1 | 0 | 1 |
| Hidupku takkan sama(28) | 0 | 0 | 0 |
| Hujan Turun(30) | 0 | 0 | 0 |
| Jadikan Kami Setia(34) | 0 | 0 | 0 |
| Kita(40) | 1 | 0 | 1 |
| Lonteku(44) | 1 | 0 | 1 |
| Melompat Lebih Tinggi(46) | 1 | 0 | 1 |
| Menyenangkan hatimu(47) | 0 | 0 | 0 |
| Pemuja Rahasia(50) | 1 | 0 | 1 |
| Pertolonganmu(52) | 0 | 0 | 0 |
| SeberapaPantas(58) | 1 | 0 | 1 |
| Sephia(59) | 1 | 0 | 1 |
| Serdadu(60) | 1 | 0 | 1 |
| Seribu Lilin(61) | 0 | 0 | 0 |
| Sungai sukacitaMu(63) | 0 | 0 | 0 |
| Tuhanlah Gembalaku(68) | 0 | 0 | 0 |
| Tuhanlah perlindunganku(69) | 0 | 0 | 0 |
| Ujung Aspal Pondok Gede(72) | 1 | 0 | 1 |
| Aku Bukan Pilihan(1) | 1 | 1 | 0 |
| Allah itu baik(3) | 0 | 1 | 1 |
| Allah Sanggup(5) | 0 | 1 | 1 |
| Anugrah Terindah(6) | 1 | 1 | 0 |
| Bahasa kasih(7) | 0 | 1 | 1 |
| Bapa Engkau Sungguh Baik(8) | 0 | 1 | 1 |
| Bento(12) | 1 | 1 | 0 |
| Berhenti Berharap(13) | 1 | 1 | 0 |
| Betapa(15) | 1 | 1 | 0 |
| Bila Engkau tak BesertaKu(16) | 0 | 1 | 1 |
| Bila Kau Tak Disampingku(17) | 1 | 1 | 0 |
| Bongkar(18) | 1 | 1 | 0 |
| Buat Aku Tersenyum(19) | 1 | 1 | 0 |
| Buku Ini Aku Pinjam(20) | 1 | 1 | 0 |
| Dan(22) | 1 | 1 | 0 |
| Film Favorit(25) | 1 | 1 | 0 |
| Hari Bersamanya(26) | 1 | 1 | 0 |
| Hari yang terindah(27) | 1 | 1 | 0 |
| Hingga Ujung Waktu(29) | 1 | 1 | 0 |
| Ibu(31) | 1 | 1 | 0 |
| Ijinkan Aku Menyayangimu(32) | 1 | 1 | 0 |
| Itu Aku(33) | 1 | 1 | 0 |
| JAP(35) | 1 | 1 | 0 |
| Kasih Persaudaraan(36) | 0 | 1 | 1 |
| Kaulah Harapan(37) | 0 | 1 | 1 |
| Kemesraan(38) | 1 | 1 | 0 |
| Kisah Klasik(39) | 1 | 1 | 0 |
| Kumenanti Seorang Kekasih(41) | 1 | 1 | 0 |
| Lapang Dada1(42) | 1 | 1 | 0 |
| Lihat Dengar Rasakan(43) | 1 | 1 | 0 |
| Manusia Setengah Dewa(45) | 1 | 1 | 0 |
| Mudah Saja(48) | 1 | 1 | 0 |
| Pasti Ku Bisa(49) | 1 | 1 | 0 |
| Perhatikan Rani(51) | 1 | 1 | 0 |
| Pesawat Tempurku(53) | 1 | 1 | 0 |
| Pria Kesepian(54) | 1 | 1 | 0 |
| Saat Aku Lanjut Usia(55) | 1 | 1 | 0 |
| Sahabat Sejati(56) | 1 | 1 | 0 |
| Sarjana Muda(57) | 1 | 1 | 0 |
| Sumbang(62) | 1 | 1 | 0 |
| Surat Buat Wakil Rakyat(64) | 1 | 1 | 0 |
| Takkan pernah terlambat(65) | 0 | 1 | 1 |
| Terlalu Singkat(66) | 1 | 1 | 0 |
| Tikus Tikus Kantor(67) | 1 | 1 | 0 |
| Tunggu Aku di Jakarta(70) | 1 | 1 | 0 |
| Tunjuk Satu Bintang(71) | 1 | 1 | 0 |
| Yang Terlewatkan(73) | 1 | 1 | 0 |
| Yang Terlupakan(74) | 1 | 1 | 0 |
|  |  |  | 18 |

Hitung akurasi = jumlah error / jumlah data.

Error = 18/74\*100%= 24%

Akurasi= 56/74\*100%= 76%

1. Kesimpulan

Buatlah simpulan akhir aplikasi yang anda kembangkan.

Menurut kami Aplikasi ini bisa dibilang cukup bagus karena memiliki akurasi 76% yang mana seharusnya aplikasi yang baik memiliki akurasi diatas >= 90%.

1. Daftar Pustaka (standar MLA)

* Daftar Pustaka

[https://lirik.kapanlagi.com](https://lirik.kapanlagi.com/)

<https://dataq.wordpress.com/2013/06/16/perbedaan-precision-recall-accuracy/>

<https://elib.unikom.ac.id/download.php?id=184459>

<https://journal.unnes.ac.id/nju/index.php/jte/article/view/10955>

* Link Github

<https://github.com/egaprasetyo/dokumenRetrieval>

* Link Youtube

<https://youtu.be/hUugttbTZ1E>