

# ANALISIS DAN IMPLEMENTASI ALGORITMA DAMERAU LEVENSHTTEIN DISTANCE UNTUK CONTENT BASED MUSIC RETRIEVAL ANALYSIS AND IMPLEMENTATION DAMERAU LEVENSHTTEIN DISTANCE ALGORITHM FOR CONTENT BASED MUSIC RETRIEVAL

Rika Dwi Erisandi<sup>1</sup>, Z.k. Abdurahman Baizal<sup>2</sup>, Tjokorda Agung Budi Wirayuda<sup>3</sup>

<sup>1</sup>Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

---

## Abstrak

Proses pencarian file musik menjadi suatu hal yang menarik apabila berdasarkan content pada input querynya. Pencarian berdasar content ini menggunakan melodi untuk mencari musik yang diinginkan. File musik yang digunakan berformat midi. Dimana midi merupakan representasi digital musik yang sederhana. File midi tersebut diubah ke dalam bentuk teks menggunakan tools midi2text.exe. Setelah didapat dalam bentuk teks kemudian proses ekstraksi, standarisasi, dan pencocokan melodi dilakukan.

Proses pencocokan melodi menggunakan algoritma Damerau-Levenshtein distance. Algoritma ini merupakan algoritma yang merepresentasikan fungsi pengukuran jarak yang mengembalikan jarak minimumnya dimana operasi deletion, insertion, substitution dan transposisi dilakukan pada dua karakter string yang dibandingkan. Damerau-Levenshtein distance merupakan pengembangan dari algoritma Levenshtein distance. Levenshtein distance juga merupakan fungsi pengukuran jarak yang mengembalikan jarak minimumnya dimana proses deletion, insertion dan substitution dilakukan dalam proses pencocokan string.

Berdasarkan hasil pengujian, kedua algoritma tersebut dapat menemukan file musik yang dicari dengan persen jarak yang sama. Dari segi performansi waktu pemrosesan query, algoritma Damerau-Levenshtein distance lebih lama  $\pm 1.906$  detik daripada algoritma Levenshtein distance untuk semua jenis query dengan output rata-rata hasil pencarian yang relatif sama banyaknya.

**Kata Kunci :** conten musik, Damerau-Levenshtein distance, Levenshtein distance,

---

## Abstract

Searching process for music file becomes interesting if based on the content of the query input. This searching, that based on the content, uses melody to search the music that we want. We use music files with midi format. Where, midi is a simple digital music representation. This midi file is changed to text type using midi2text.exe tools. After we get the text type, then we continue with extraction, standardization, and melody match process.

We do melody match process using Damerau-Levenshtein distance algorithm. This algorithm is an algorithm that represent distance measuring function that return the minimum distance where deletion, insertion, substitution, and transposition process are done in two string character that we compare. Damerau-Levenshtein distance is the development of Levenshtein distance algorithm. Levenshtein distance is also a distance measuring function that return the minimum distance where deletion, insertion, and substitution process are done in the string match process. There is no transposition process in Levenshtein distance algorithm.

Based on the experiment result, both of the algorithm above can find the music file that we search with equal percent distance. From time performance of query processing, Damerau-Levenshtein distance algorithm is  $\pm 1.906$  seconds longer than Levenshtein distance algorithm for all query type with the same relative amount of the average output from searching result.

**Keywords :** music content, Damerau-Levenshtein distance, Levenshtein distance,

---

# 1. PENDAHULUAN

## 1.1 Latar Belakang Masalah

Seiring dengan berkembangnya teknologi, setiap orang dapat dengan mudah mendapatkan file musik dari internet dalam berbagai jenis, ukuran, maupun format. Hal ini membuat jumlah file musik semakin bertambah banyak. Permasalahan yang muncul kemudian adalah bagaimana mencari file musik yang diinginkan. Secara umum, pencarian musik biasanya berdasarkan judul lagu atau penyanyinya disebut dengan pencarian berdasarkan bentuk metadata. Akan tetapi, pencarian file musik juga dapat dilakukan berdasarkan *content*-nya disebut dengan *content based music retrieval*.

Pencarian musik berdasarkan *content* ini menggunakan salah satu konsep pencocokan string yaitu *approximate string matching*. Teknik ini akan membandingkan dua berkas file musik dalam format MIDI yang diubah ke dalam bentuk teks. Algoritma yang akan digunakan pada proses pencocokan yaitu *Levenshtein Distance* dan *Damerau-Levenshtein Distance*. Algoritma *Levenshtein Distance* melakukan proses pencocokan dengan cara menghitung jumlah minimum operasi *insertion*, *deletion* dan *substitution* terhadap karakter suatu string [5]. Performansi dan akurasi algoritma *Levenshtein Distance* dalam proses pencocokan masih dapat ditingkatkan dengan menggunakan algoritma *Damerau-Levenshtein Distance*. Algoritma *Damerau-Levenshtein Distance* menggunakan operasi *insertion*, *deletion*, *substitution* dan *transposisi* dalam proses pencocokan terhadap karakter suatu string [2].

Berdasarkan hal-hal yang dikemukakan di atas, muncul ide untuk mengimplementasikan algoritma *Damerau-Levenshtein Distance* untuk *content based music retrieval* sebagai pengembangan dari algoritma *Levenshtein Distance*.

## 1.2 Perumusan Masalah

Berdasarkan latar belakang di atas, berikut rumusan masalah yang akan diteliti:

- Bagaimana mengimplementasikan *Content Based Music Retrieval* dengan menggunakan algoritma *damerau-levenshtein distance*?
- Bagaimana membandingkan performansi antara kedua algoritma dilihat dari output file musik yang ditemukan?
- Bagaimana mengukur peningkatan performansi algoritma *damerau-levenshtein distance* dibandingkan dengan *levenshtein distance*?

Adapun batasan dari masalah yang akan diteliti adalah sebagai berikut:

- File musik yang digunakan dalam format MIDI (*Musical Instrument Digital Interface*). Format MIDI merupakan format musik paling dasar yang menggambarkan informasi *content* musik.
- Digunakan file MIDI format 1, terdiri dari beberapa *track* yang berjalan secara bersamaan.

- Inputan yang digunakan adalah potongan melodi MIDI :
  - Potongan melodi pendek yang dimainkan berdurasi  $\pm 30$  detik.
  - Potongan melodi sedang yang dimainkan berdurasi  $\pm 1$  menit.
  - Potongan melodi panjang yang dimainkan berdurasi  $\geq 2$  menit.
- Pencarian musik hanya terbatas pada fitur-fitur dalam musik MIDI berupa melodi tidak menambahkan dengan fitur musik yang lain.

### 1.3 Tujuan

Tujuan penyusunan tugas akhir ini sebagai berikut:

- Mengimplementasikan metode *approximate string matching* dengan algoritma *Damerau-Levenshtein Distance* untuk *Content Based Music Retrieval*.
- Menganalisis performansi metode *approximate string matching* dengan algoritma *Damerau-Levenshtein Distance* yang diukur dengan parameter :
  - Ketepatan dalam menemukan file musik yang dicari dalam database, diwakili dengan parameter  $d$  (*distance*) yang diubah ke dalam persentase nilai. Dimana semakin minimum jarak maka ketepatannya semakin tinggi.
  - Kecepatan dalam menemukan file musik yang dicari dalam database, diwakili dengan parameter waktu.
  - Jumlah hasil output file musik terhadap panjang file musik yang diinputkan (panjang melodi dalam durasi waktu).

### 1.4 Metodologi Penyelesaian Masalah

Metodologi yang digunakan dalam penyusunan tugas akhir ini adalah:

1. Studi Literatur  
Melakukan pengumpulan dan pembelajaran materi yang berhubungan dengan penyusunan tugas akhir tentang algoritma *damerau-levenshtein distance*, algoritma *levenshtein distance*, format MIDI dan *content based music retrieval*. Literatur ini bersumber dari paper dan referensi lainnya dari internet.
2. Pengumpulan Data  
Melakukan pengumpulan file musik midi yang akan dipergunakan dalam perangkat lunak.
3. Analisis dan Perancangan  
Tahap *preprocessing* meliputi membaca file midi setelah diubah ke dalam bentuk teks, ekstraksi melodi, standarisasi melodi, kemudian proses pencocokan query. Analisis perancangan menggunakan Data Flow Diagram (DFD) yaitu diagram yang menggambarkan aliran data dari suatu sistem, perancangan *user interface* dan perancangan basis data.

4. Implementasi  
Implementasi menggunakan Borland Delphi 7.0, database menggunakan Oracle 10g, dan midi2txt.exe sebagai *tools* yang membantu pembacaan informasi dalam file MIDI.
5. Pengujian  
Pengujian meliputi pengujian fungsionalitas untuk menguji apakah fungsi yang dihasilkan telah sesuai dengan keinginan pengguna dan pengujian pemrosesan query. Hasil pengujian pemrosesan query akan dianalisis dan diambil kesimpulan.
6. Penyusunan Laporan  
Menyusun laporan hasil penelitian berupa buku Tugas Akhir.



## 5. Kesimpulan dan Saran

### 5.1 Kesimpulan

Kesimpulan yang dapat ditarik dari tugas akhir ini yaitu:

1. Dilihat dari contoh hasil pemrosesan query, algoritma Damerau-Levenshtein distance maupun Levenshtein distance dapat menemukan file musik yang dicari.
2. Dari segi performansi waktu pemrosesan query untuk semua jenis query (query pendek, sedang, panjang) proses query menggunakan algoritma Damerau-Levenshtein distance lebih lama daripada Levenshtein distance sebesar  $\pm 1.906$  detik. Hal tersebut terjadi karena pada Damerau-Levenshtein distance dilakukan pengecekan kembali terhadap suatu karakter string yang dibandingkan apakah memenuhi kondisi transposisi, hal tersebut yang membuat proses query menggunakan Damerau-Levenshtein distance lebih lama daripada Levenshtein distance.
3. Algoritma Damerau-Levenshtein distance dan Levenshtein distance menghasilkan jumlah output rata-rata hasil pencarian yang relatif sama banyaknya untuk semua jenis query (pendek, sedang, panjang).
4. Dilihat dari koefisien korelasi dan determinasi pada proses query menggunakan algoritma Damerau-Levenshtein distance dapat diambil kesimpulan bahwa antara variabel output hasil dengan waktu eksekusi pemrosesan query menghasilkan hubungan tanpa korelasi dan kurang memiliki keterhubungan secara fungsional. Hal ini karena proses pencocokannya dilakukan terhadap seluruh data yang ada di database.
5. Secara subjektif, bahwa user menganggap lagu yang dijadikan sebagai input query terdengar sama dengan hasil query, walaupun lagu sebagai input query dimodifikasi.

### 5.2 Saran

Adapun saran yang dapat disampaikan untuk proses pengembangan lebih lanjut adalah:

1. Menggunakan *query by humming* yaitu query menggunakan suara manusia sebagai inputannya yang melantunkan melodi dari musik yang akan dicari.
2. Format midi yang digunakan yaitu format 0 dan 2 selain format 1.
3. Algoritma yang digunakan pada proses query yaitu *bitap algorithm* merupakan pengembangan dari algoritma Damerau-Levenshtein distance.

## Daftar Pustaka

- [1] Chen, Arbe.L.P, Music Retrieval and Analysis, 2003, National Tsing Hua University, [www.make.cs.nthu.edu.tw/alp/ismir03\\_tutorial\\_new.ppt](http://www.make.cs.nthu.edu.tw/alp/ismir03_tutorial_new.ppt), di download pada bulan Maret 2008
- [2] Damerau-Levenshtein distance, [http://en.wikibooks.org/wiki/Algorithm\\_implementation/Strings/Levenshtein\\_distance](http://en.wikibooks.org/wiki/Algorithm_implementation/Strings/Levenshtein_distance) di download pada bulan Januari 2008
- [3] Interval, <http://id.wikipedia.org/wiki/Interval>, di download pada bulan Februari 2008
- [4] Korelasi, <http://wikipedia.org/wiki/korelasi> , di download pada bulan Agustus 2008
- [5] Levenshtein distance, [http://en.wikibooks.org/wiki/Algorithm\\_implementation/Strings/Levenshtein\\_distance](http://en.wikibooks.org/wiki/Algorithm_implementation/Strings/Levenshtein_distance) di download pada bulan Januari 2008
- [6] Modul Praktikum Analisis Data, 2004, Laboratorium Matematika dan Statistika, Bandung , Sekolah Tinggi Teknologi Telkom
- [7] Modular Arithmetic, [http://en.wikipedia.org/wiki/Modular\\_arithmetic](http://en.wikipedia.org/wiki/Modular_arithmetic), di download Maret 2008
- [8] Multimedia Information Retrieval: Midi as a format for content based retrieval of audio, [www.computing.dcu.ie/~asmeaton/pubs/Midi-retrieval](http://www.computing.dcu.ie/~asmeaton/pubs/Midi-retrieval), di download pada bulan Februari 2008
- [9] Musical Instrument Digital Interface, [http://en.wikipedia.org/wiki/Musical\\_Instrument\\_Digital\\_Interface](http://en.wikipedia.org/wiki/Musical_Instrument_Digital_Interface) di download bulan Februari 2008
- [10] Nada, <http://id.wikipedia.org/wiki/Nada>, di download pada bulan Februari 2008
- [11] Outline of the standard MIDI file structure, [craig@ccrma.stanford.edu http://www.ccarh.org/courses/253/handout/smf/](http://www.ccarh.org/courses/253/handout/smf/) di download pada bulan Februari 2008

- [12] Sandra Indriati, Susan, 2007, Tugas Akhir: Content Based Music Retrieval Dengan Teknik Approximate String Matching, STT Telkom, Bandung
- [13] Semua istilah tentang musik, <http://musikpedia.com> di download bulan Februari 2008
- [14] Tangga nada, [http://id.wikipedia.org/wiki/Tangga\\_nada](http://id.wikipedia.org/wiki/Tangga_nada), di download pada bulan Februari 2008
- [15] Teori Time Modul, <http://www.musicmarkup.info/modules/time/index.html>, didownload pada tanggal Maret 2008
- [16] Uitdenbogerd Alexandra, Zobel Justin, Matching Technique for Large Music Databases, Departement of Computer Science, RMIT University Australia, [www.cs.rmit.edu.au/~jz/fulltext/acm-mm99.pdf](http://www.cs.rmit.edu.au/~jz/fulltext/acm-mm99.pdf), didownload pada bulan Februari 2008
- [17] Veltkamp Remco, Mulimedia Retrieval : Music Retrieval, [www.cs.uu.nl/docs/vakken/mr/coursematerial/mr-music.ppt](http://www.cs.uu.nl/docs/vakken/mr/coursematerial/mr-music.ppt), didownload pada Februari 2008