**JURUSAN TEKNIK INFORMATIKA**

**FAKULTAS TEKNOLOGI INFORMASI**

**INSTITUT TEKNOLOGI SEPULUH NOPEMBER**

**USULAN TUGAS AKHIR**

# **IDENTITAS PENGUSUL**

Nama : **I Wayan Surya Priantara**

NRP : **5107 100 057**

Dosen Wali : **Ir. Muchammad Husni, M.Kom**

1. **JUDUL TUGAS AKHIR**

***“Implementasi Deteksi Penjiplakan dengan Algoritma Winnowing pada Dokumen Terkelompok“***

1. **LATAR BELAKANG**

Plagiat atau penjiplakan, merupakan kegiatan menduplikasi sebagian atau seluruh karya orang lain tanpa mencantumkan narasumber. Dalam bidang pendidikan kegiatan ini sering dilakukan oleh pelajar dalam pembuatan tugas seperti tugas sekolah/kuliah dan pembuatan laporan. Di sisi pengajar, kegiatan penjiplakan ini sangat menyusahkan dalam hal evaluasi dan penilaian, walau sudah ada sanksi yang tegas jika ada seorang pelajar melakukan penjiplakan. Kegiatan penjiplakan susah dideteksi bila dokumen tugas yang diperiksa berjumlah sangat banyak dan seorang pengajar mengajar lebih dari satu mata pelajaran.

Kegiatan deteksi penjiplakan ini dilakukan untuk mengetahui tingkat penjiplakan suatu dokumen pada dokumen lain. Sehingga, dapat mengetahui tingkat atau presentase seseorang melakukan penjiplakan pada suatu dokumen. Selain itu dapat memudahkan pekerjaan seorang pengajar dalam hal mengevaluasi tugas-tugas yang berbentuk dokumen.

Saat ini sudah terdapat sistem yang mampu mendeteksi penjiplakan pada beberapa dokumen, untuk membantu pengajar menyelesaikan pekerjaannya. Pada sistem ini mampu melakukan pengecekan yang bersifat *one to many* hingga *many to many*. Sehingga dapat digunakan pada dokumen-dokumen yang berjumlah banyak

Namun, pada sistem tersebut memiliki waktu komputasi yang cukup lama ketika melakukan pengecekan penjiplakan pada dokumen yang berjumlah banyak. Hal ini di karenakan sistem tersebut membandingkan satu persatu dokumen yang ada, baik dokumen yang memiliki topik yang sama maupun berbeda topik. Oleh karena itu diperlu sistem yang dapat mengelompokan secara otomatis dokumen-dokumen yang memiliki topik yang sama. Diharapkan pengelompokan secara otomatis dokumen-dokumen ini menyebabkan waktu proses dalam pengecekan penjiplakan menjadi lebih singkat.

1. **TUJUAN TUGAS AKHIR**

Membangun sistem yang dapat melakukan pengecekan penjiplakan pada dokumen-dokumen yang berjumlah banyak. Dokumen yang bejumlah banyak ini menyebabkan waktu komputasi yang lama, sehingga diperlukan proses pengelompokan secara otomatis. Dengan pengelompokan ini maka waktu komputasi akan menjadi lebih singkat

1. **PERMASALAHAN**

Permasalahan yang diangkat dalam Tugas Akhir ini dapat dipaparkan sebagai berikut:

1. Bagaimana pengelompokan dokumen-dokumen yang sudah menjadi bentuk fingerprint
2. Bagaimana penentuan centroid dalam pengelompokan dokumen
3. Berapa jumlah kelompok yang cocok pada dokumen-dokumen yang akan dikelompokan
4. **BATASAN MASALAH**

Permasalahan yang dibahas dalam Tugas Akhir ini memiliki batasan-batasan seperti dataset yang akan dilakukan pendeteksian adalah file teks digital seperti .doc. Dokumen-dokumen yang akan dideteksi berdasarkan susunan-susunan kata pada tiap-tiap dokumen. Untuk pengelompokan dokumen-dokumen digunakan pengelopompokan yang bersifat partisi

1. **RINGKASAN TUGAS AKHIR**

Pada tugas akhir ini, akan diimplementasikan algoritma winnowing dalam pengecekan penjiplakan pada dokumen terkelompok. Dokumen-dokumen yang di kelompokan ini bertujuan untuk mempercepat waktu pengecekan penjiplakan pada dokumen.



Gambar . Alur Proses

Input dari sistem ini adalah dokumen-dokumen yang akan dilakukan pengecekan penjiplakan. Untuk melakukan pengecekan penjiplakan terdapat kebutuhan mendasar yang harus dipenuhi oleh suatu algoritma penjiplakan seperti

1. *Whitespace Insensitivity*, yang berarti dalam melakukan pencocokan terhadap file teks seharusnya tidak terpengaruh oleh spasi, jenis huruf (kapital atau normal), tanda baca dan sebagainya.
2. *Noise Surpression,* yang berarti menghindari penemuan kecocokan dengan panjang kata yang terlalu kecil atau kurang relevan, misal: ‘the’. Panjang kata yang ditengarai merupakan penjiplakan harus cukup untuk membuktikan bahwa kata-kata tersebut telah dijiplak dan bukan merupakan kata yang umum digunakan.
3. *Position Independence*, yang berarti penemuan kecocokan / kesamaan tidak harus bergantung pada posisi kata-kata. Walau tidak dalam berada posisi yang sama pencocokan juga harus dilakukan.

Pada tahap selanjutnya akan dilakukan pembentukan rangkaian gram dari teks yang telah dibersihkan dari *Whitespace Insensitivity, Noise Surpression, Position Independence* dengan ukuran gram yang telah ditentukan[2].

Pada tahap selanjutnya dilakukan perhitungan nilai-nilai *hash* dari setiap gram menggunakan *rolling hash*. Untuk mendapatkan nilai hash dengan menggunakan persamaan

c1\*b(k-1)+ c2\*b(k-2)+...+c(k-1)\*bk+ck (1)

dimana :

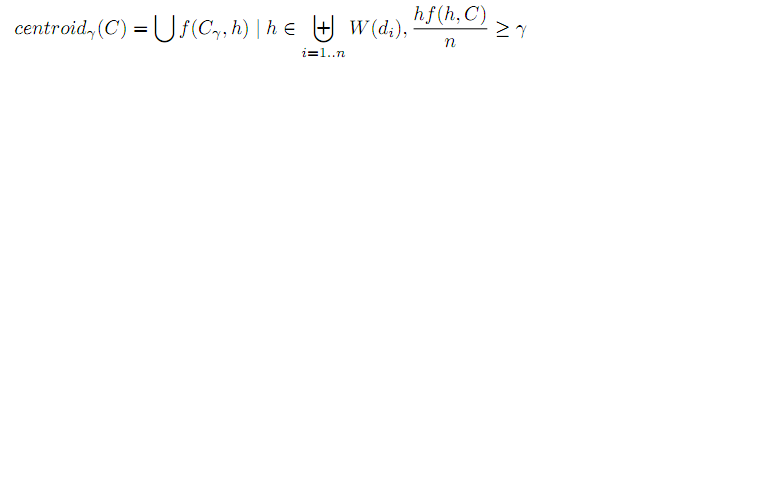
c : ascii karakter

b : bilangan prima

k : banyak karakter

Dari persamaan tersebut akan di dapatkan nilai-nilai hash tiap-tiap dokumen. Dari nilai-nilai hash tiap dokumen selanjutnya dibentuk window dengan ukuran window yang ditentukan. Dari window-window yang telah dibentuk selanjut dipilih nilai terkecil dari masing-masing window, Bila nilai hash yang terpilih kembali pada window selanjutnya maka nilai tersebut tidak di ambil. Tiap-tiap dokumen memiliki nilai hashing (W(d)) dan panjang yang berbeda.

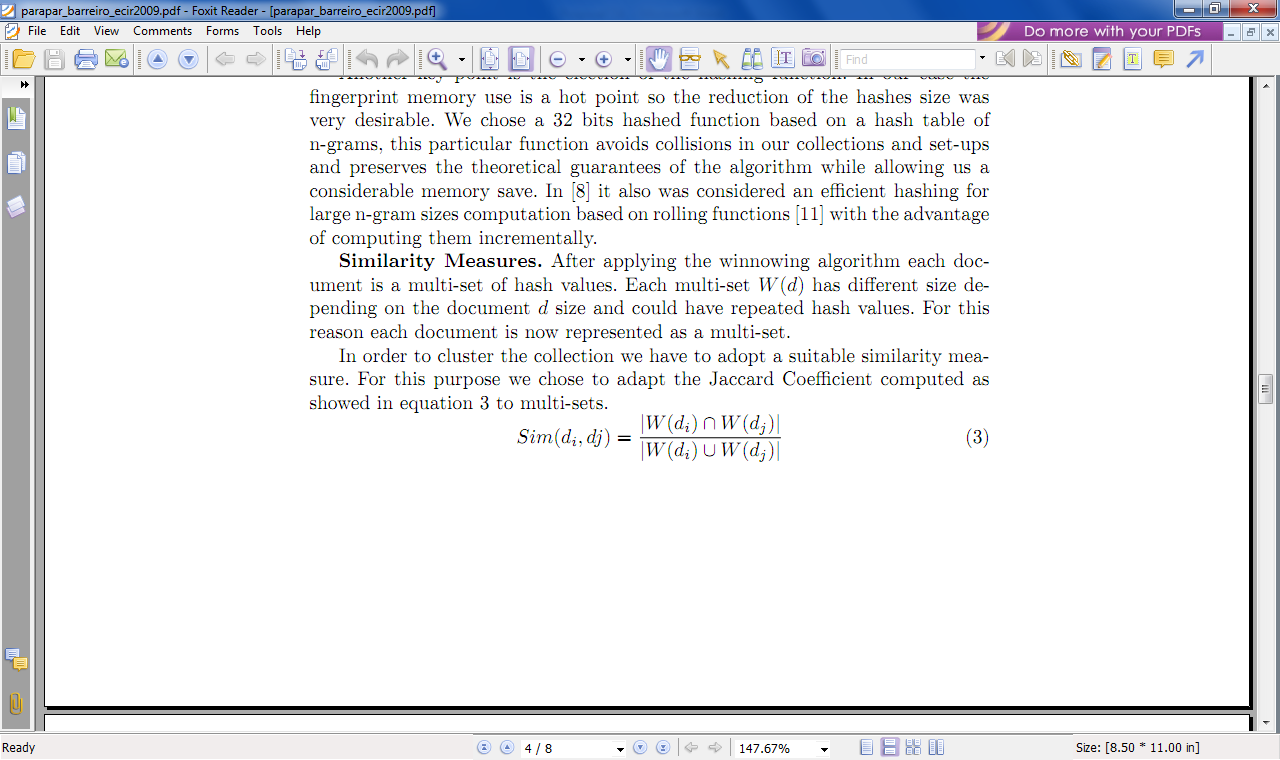
Tahap selanjutnya dilakukan pengelompokan dokumen-dokumen berdasarkan kemiripan karakteristik dari tiap-tiap dokumen. Untuk penentuan centroid (C) dengan multiset nilai-nilai hashing dapat menggunakan persamaan[1]



(2)

dimana f(Cᵧ, h) merupakan fungsi yang mengembalikan nilai hashing(h) sebanyak sedangkan hf(h, C) merupakan nilai kemunculan h pada gabungan dokumen tiap cluster.

Setelah terbentuk kelompok-kelompok dokumen selanjutnya akan dilakukan pengecekan kesamaan dokumen per kelompok. Untuk dapat menghitung kesamaan antar dokumen dapat menggunakan persamaan



(3)

dari persamaan tersebut akan menghasilkan nilai kemiripan antar dokumen. Hasil proses ini selanjutnya akan di tampilkan kehadapan user.

1. **METODOLOGI**

Ada beberapa tahap dalam proses pengerjaan tugas akhir ini. Berikut tahap-tahap dalam pembuatannya :

1. **Studi literatur**

Pada tahap ini akan dipelajari sejumlah literatur mengenai konsep dan teknologi yang akan digunakan. Literatur yang digunakan meliputi paper referensi, buku referensi, ebook dan dokumentasi internet, diantaranya tentang penggunaan algoritma winnowing dalam pengecekan penjiplakan, algoritma hashing dalam pembentukan nilai-nilai hash pada dokumen dan algoritma clustering dalam pengelompokan dokumen yang sudah dalam bentuk hash.

1. **Perancangan perangkat perangkat lunak dan desain sistem**

Pada tahap ini dilakukan proses perancangan perangkat lunak, berdasarkan literatur yang telah di kaji kemudian di buat desain model data, diagram alir proses-proses yang ada, dan desain antar muka aplikasi. Yang selanjutnya kemudian akan diimplemetasikan.

1. **Pengimplementasian perangkat lunak**

Pada tahap ini dilakukan proses pengimplemetasian, dengan berdasar pada rancangan awal perangkat lunak dan literature yang telah dikaji. Aplikasi telah mulai dibuat secara menyeluruh.

1. **Uji coba dan evaluasi**

Pada tahap ini akan diuji aplikasi yang telah selesai diimplemetasikan tersebut. Pengujian dan evaluasi akan dilakukan pada dokumen-dokumen.

1. **Penyusunan laporan tugas akhir**

Pada tahap ini disusun buku sebagai dokumentasi dari pelaksanaan tugas akhir, yang mencakup seluruh konsep, teori, implementasi, serta hasil yang telah dikerjakan. Laporan tugas akhir ini akan dibagi menjadi beberapa bab sebagai berikut:

1. Bab I, Pendahuluan, berisi latar belakang, permasalahan, tujuan, batasan permasalahan, metodologi, dan sistematika penulisan.
2. Bab II, Landasan Teori, akan dibahas dasar ilmu yang mendukung pembahasan tugas akhir ini.
3. Bab III, Desain Aplikasi.
4. Bab IV, Implementasi dari aplikasi yang telah dibuat, akan dilakukan pembuatan aplikasi yang dibangun dengan komponen-komponen yang telah ada yang sesuai dengan permasalahan dan batasannya yang telah dijabarkan pada bab pertama.
5. Bab V, Uji coba dan analisa hasil, akan dilakukan uji coba berdasarkan parameter-parameter yang ditetapkan, dan kemudian dilakukan analisa terhadap hasil uji coba tersebut.
6. Bab VI, Penutup, berisi kesimpulan yang dapat diambil dari Tugas Akhir ini beserta saran untuk pengembangan selanjutnya
7. **DAFTAR PUSTAKA**

[1] Javier Parapar and Alvaro Barreiro, *Evaluation of Text Clustering Algorithms with N-Gram-Based Document Fingerprints*, IRLab, Computer Science Department University of A Coruna, Spain

[2] Schleimer, S., Wilkerson, D., dan Aiken, A. 2003. *Winnowing: Local algorithms for document fingerprinting*. In Proceedings of the ACM SIGMOD international conference on management of data. pp 76–85.

1. **JADWAL PEMBUATAN TUGAS AKHIR**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Kegiatan** | **Bulan** | | | | | | | |
| **September** | | **Oktober** | | **November** | | **Desember** | |
| Studi Literatur |  |  |  |  |  |  |  |  |
| Perancangan Perangkat Lunak dan Desain Sistem |  |  |  |  |  |  |  |  |
| Pengimplementasian Perangkat Lunak |  |  |  |  |  |  |  |  |
| Uji Coba dan Evaluasi |  |  |  |  |  |  |  |  |
| Penyusunan Laporan Tugas Akhir |  |  |  |  |  |  |  |  |

**LEMBAR PENGESAHAN**

###### **Surabaya, 29 Oktober 2010**

Menyetujui,

Dosen Pembimbing I

# **(Diana Purwitasari, S.Kom., M.Sc.)**

# **( NIP.19780410 200312 2001)**

Dosen Pembimbing II

(Umi Laili Yuhana, S.Kom, M.Sc.)

( NIP.19790626 200501 2002)