

PENDETEKSIAN PLAGIARISME MENGGUNAKAN ALGORITMA RABIN-KARP DENGAN METODE ROLLING HASH

Joko Priambodo

Teknik Informatika, Universitas Pamulang

e-mail: dosen00276@unpam.ac.id

ABSTRAK

Plagiarisme adalah tindakan penyalahgunaan, pencurian/perampasan, penerbitan, pernyataan, atau menyatakan sebagai milik sendiri sebuah pikiran, ide, tulisan, atau ciptaan yang sebenarnya milik orang lain. Plagiat atau biasa disebut penjiplakan adalah sebuah masalah yang cukup signifikan pada akademisi di perguruan tinggi. Hal plagiat yang biasanya dilakukan terhadap konten digital adalah melakukan *copy-paste*, *quote*, dan revisi terhadap dokumen asli. Untuk mengantisipasinya, dibutuhkan suatu cara yang dapat menganalisis teknik-teknik plagiat yang dilakukan. Ada beberapa pendekatan yang bisa diambil, salah satunya dengan menggunakan algoritma *Rabin-Karp* dengan metode *Rolling Hash*. Pendeteksian plagiarisme menggunakan algoritma *Rabin-Karp* dengan metode *rolling hash* ini diimplementasikan ke dalam program atau aplikasi untuk menentukan nilai tingkat akurasi dengan nilai presentase.

Kata kunci: Plagiarisme, Algoritma Rabin-Karp, *Rolling Hash*.

1. LATAR BELAKANG

Dokumen digital memang dapat dijiplak dengan sangat mudah. Praktik plagiat atau penjiplakan ini sudah sering terjadi khususnya pada kalangan akademisi baik lingkungan sekolah maupun perguruan tinggi. Tindakan plagiat yang dilakukan siswa atau mahasiswa ini sangat tidak mencerminkan sikap kreatif dan terpelajar sebagai kaum intelektual. Bentuk plagiat ini dapat bermacam-macam seperti misalnya yang paling sering dilakukan yaitu *copy-paste-edit* suatu dokumen baik merupakan pekerjaan teman maupun dokumen yang berasal dari *website*, sehingga menjadi suatu hasil yang berbeda dari sebelumnya. Selain itu plagiat dapat pula ditemukan dalam bentuk *quote* atau kutipan pada sebuah dokumen.

Dalam mendeteksi plagiat secara manual, hal terpenting yang harus dilakukan adalah mengetahui bodi inti dari suatu dokumen. Metode berikutnya adalah membuat katalog *paperwork* dan dokumen-dokumen terdahulu. Hal ini biasanya sudah dilakukan pada institusi besar terhadap hasil kerja para mahasiswanya. Selain itu seorang dosen misalnya, dapat pula melakukan pengecekan terhadap gaya penulisan dokumen mahasiswanya. Akan tetapi, seluruh jenis deteksi manual memiliki masalah defisiensi yang cukup serius. Sangat tidak mungkin untuk memeriksa sebuah dokumen dengan ribuan dokumen lainnya yang terdapat dalam katalog, apalagi menafsirkan gaya penulisannya.

Penggunaan *search engine* dapat membantu melakukan pencarian, yaitu dengan memasukkan kata kunci tema dokumen dan biarkan *search engine* menemukan dokumen yang dijiplak yang cocok. Ini memang berguna apabila keseluruhan dokumen yang dijiplak. Namun kurang efektif apabila plagiator hanya menggunakan sebagian dari artikel atau menggabungkan beberapa pecahan artikel. Selain itu cara ini juga banyak mengkonsumsi waktu.

Metode terakhir adalah dengan cara melakukan perbandingan dengan sumber dokumen asli. Sejumlah dokumen yang dapat dijadikan sumber dokumen asli dapat merupakan *input* secara manual maupun dokumen yang tersebar lewat internet. Pada umumnya, pada metode inilah algoritma *Rabin-Karp* dan variasinya diterapkan untuk mendapatkan hasil yang maksimal. Algoritma ini biasanya diimplementasikan pada sebuah aplikasi baik yang berbasis *desktop* maupun yang berbasis *web*.

Algoritma Rabin-Karp adalah suatu algoritma pencarian string yang ditemukan oleh Michael Rabin dan Richard Karp. Algoritma ini menggunakan *hashing* untuk menemukan sebuah *substring* dalam sebuah teks. *Hashing* adalah metode yang menggunakan fungsi *hash* untuk mengubah suatu jenis data menjadi beberapa bilangan bulat sederhana. Disebut algoritma “pencarian string” dan bukan “pencocokan string” seperti Knuth-Morris-Pratt atau Boyer-

Moore karena memang algoritma Rabin-Karp tidak bertujuan menemukan string yang cocok dengan string masukan, melainkan menemukan pola (pattern) yang sekiranya sesuai dengan teks masukan.

Plagiarisme selalu menjadi perhatian di banyak sektor, terutama sektor akademis. Sering kita temui pekerjaan teman yang dihargai dengan nilai “0” karena diketahui telah menyalin hasil pekerjaan temannya yang lain. Praktik menyalin beberapa bagian atau keseluruhan tulisan tanpa mencantumkan sumber menjadi hal biasa yang sering ditemukan pada tugas harian, makalah, bahkan skripsi. Pencegahan dan pendeteksian merupakan cara yang dapat dilakukan untuk mengurangi plagiarisme. Pencegahan berarti menghalangi munculnya plagiarisme yang lebih ditekankan kepada moral masyarakat dan sistem pendidikan. Cara ini akan memberikan efek jangka panjang yang luar biasa. Pendeteksian berarti cara untuk mengungkapkan plagiarisme (Bagus H. , 2018 : p. 1).

Plagiarisme adalah tindakan penyalahgunaan, pencurian/perampasan, penerbitan, pernyataan, atau menyatakan sebagai milik sendiri sebuah pikiran, ide, tulisan, atau ciptaan yang sebenarnya milik orang lain. Plagiarisme adalah penjiplakan yang melanggar hak cipta (Pusat Bahasa Departemen Pendidikan Nasional, 2002). Hak Cipta adalah hak eksklusif bagi Pencipta atau penerima hak untuk mengumumkan atau memperbanyak ciptaannya atau memberikan izin untuk itu dengan tidak mengurangi pembatasan-pembatasan menurut peraturan perundang-undangan yang berlaku. Ciptaan adalah hasil setiap karya Pencipta yang menunjukkan keasliannya dalam lapangan ilmu pengetahuan, seni, atau sastra (Undang-Undang Republik Indonesia Nomor 19 Tahun 2002)(Nugroho E. , 2011 : p. 32).

Penjiplakan atau plagiarisme berarti mencontoh atau meniru atau mencuri tulisan dan karya orang lain yang kemudian diakui sebagai karangannya sendiri dengan ataupun tanpa seizin penulisnya. Penjiplakan dokumen digital bukanlah hal yang susah, cukup dengan menggunakan teknik copy-paste-modify pada sebagian isi dokumen dan bahkan keseluruhan isi dokumen sudah bisa dikatakan bahwa dokumen tersebut merupakan hasil duplikasi dari dokumen lain. Di Indonesia seringkali diramalkan dengan berita kasus plagiat yang dilakukan oleh beberapa akademisi di perguruan tinggi (Subekti & Fauzi, 2012 : p. 1). Pelaku yang melakukan plagiarisme memiliki beberapa alasan, alasan paling dominan

mengapa pelaku-pelaku tindak plagiat tersebut melakukan tindakan plagiarisme adalah karena mereka malas dan merasa tindakan plagiarisme adalah sebuah jalan singkat untuk menyelesaikan tugasnya. Hal ini sering terjadi di bidang akademik dan umumnya dilakukan oleh pelajar ataupun tenaga pengajar yang ingin tugas karangan atau karya ilmiah segera selesai. Tindakan plagiarisme ini bisa berdampak kepada masyarakat berupa berkurangnya kreativitas masyarakat karena akan timbulnya rasa takut karyanya dijiplak oleh orang lain, sehingga masyarakat malas berkarya dan memunculkan ide-ide baru.

Berdasarkan pernyataan di atas, penulis ingin meneliti tentang pendeteksi plagiarisme. Pada penelitian sebelumnya sudah terdapat beberapa metode untuk pendeteksi plagiarisme diantaranya: Algorithm Brute Force, Algorithm Boyer-Moore, Algorithm Rabin-Karp. Pada penelitian sebelumnya, untuk pendeteksian plagiarisme menggunakan algoritma Brute Force (Putra, 2010 : p. 1). Algoritma Brute Force memiliki kelebihan pada pencarian single pattern (satu pola), tetapi masih memiliki kelemahan pada pencarian multi pattern. Pada algoritma Boyer-Moore baik pada pergeseran dua karakter terakhir, tetapi buruk pada pergeseran sebelumnya (Singla & Garg, 2012 : p. 3).

Algoritma Rabin-Karp dapat memecahkan masalah algoritma Brute Force yaitu pada pencarian multi pattern (Singla & Garg, 2012 : p. 3) yang mana ini tepat untuk pendeteksian plagiarisme yang memerlukan pecocokan banyak pola (multi pattern). Tetapi pada algoritma Rabin-Karp masih memiliki kelemahan sulitnya keakuratan antar kata yang mirip, karena pada algoritma Rabin-Karp masih menggunakan fungsi Hash untuk mengubah kata menjadi sebuah bilangan desimal, dimana fungsi Hash suatu string S1 mungkin sama dengan string S2.

Metode Rolling Hash dapat memecahkan masalah pada fungsi hash yang terdapat di algoritma Rabin-Karp. Karena menggunakan rolling hash, nilai hash yang bisa diupdate dengan rumus tertentu berdasarkan karakter yang dibuang dan ditambahkan (Nugroho Y. , 2002 : p. 5).

Pada penelitian ini akan menerapkan algoritma Rabin-Karp dengan metode Rolling Hash untuk pendeteksi plagiarisme, sehingga hasil pendeteksian lebih akurat.

Identifikasi Masalah

- a. Masih rendahnya kesadaran dalam menghargai karya orang lain dan rendahnya kreatifitas.
- b. Algoritma Rabin-Karp masih memiliki kelemahan yaitu sulitnya mengakuratkan kata yang mirip, dimana fungsi hash yang terdapat pada algoritma Rabin-Karp memungkinkan suatu string K1 sama dengan K2.

2. METODE PENELITIAN

Metode penelitian pada dasarnya merupakan cara ilmiah untuk mendapatkan data dengan tujuan dan kegunaan tertentu. Suatu pendekatan metode penelitian digunakan untuk memecahkan masalah dalam proses penyelidikan. Metode merupakan cara seseorang dalam melakukan sesuatu untuk mencapai tujuan. Metodologi penelitian adalah penggunaan suatu cara oleh peneliti dalam pengumpulan data penelitian. Penentuan metode penelitian sangatlah penting, karena dapat membantu mengarahkan peneliti dalam mengumpulkan, mengolah dan menganalisis data penelitian.

Untuk memudahkan dalam melakukan penelitian yang berhubungan dengan masalah yang diteliti, maka penulis melakukan pendekatan studi kasus. Dengan menggunakan pendekatan ini, data yang dikumpulkan dapat disesuaikan dengan keadaan yang sebenarnya dan dibandingkan dengan teori yang menunjang. Dengan demikian, dapat memberikan gambaran yang cukup jelas serta dapat menarik kesimpulan dari objek yang diteliti.

Metode yang digunakan untuk proses dalam penelitian ini adalah sebagai berikut :

a. Wawancara

Wawancara adalah percakapan dengan maksud tertentu. Percakapan itu dilakukan oleh kedua belah pihak, yaitu pewawancara yang mengajukan pertanyaan dari yang diwawancarai yang memberikan atas pertanyaan itu. Wawancara digunakan oleh peneliti untuk menggunakan menilai keadaan seseorang. Dalam wawancara tersebut biasa dilakukan secara individu maupun dalam bentuk kelompok, sehingga didapat data informatik yang orientik.

Metode *interview* adalah sebuah dialog atau Tanya jawab yang dilakukan dua orang atau lebih yaitu pewawancara dan terwawancara (nara sumber) dilakukan secara berhadapan-hadapan.

Sedangkan *interview* yang penulis gunakan adalah jenis *interview* pendekatan

yang menggunakan petunjuk umum, yaitu mengharuskan pewawancara membuat kerangka dan garis-garis besar atau pokok-pokok yang ditanyakan dalam proses wawancara, penyusunan pokok-pokok ini dilakukan sebelum wawancara. Dalam hal ini pewawancara harus dapat menciptakan suasana yang santai tetapi serius yang artinya bahwa *interview* dilakukan dengan sungguh-sungguh, tidak main-main tetapi tidak kaku.

b. Pengamatan/Observasi

Sebagai metode ilmiah observasi dapat diartikan sebagai pengamatan, meliputi pemusatan perhatian terhadap suatu obyek dengan menggunakan seluruh alat indra. Jadi observasi merupakan suatu penyelidikan yang dilakukan secara sistematis dan sengaja diadakan dengan menggunakan alat indra terutama mata terhadap kejadian yang berlangsung dan dapat dianalisa pada waktu kejadian itu terjadi.

c. Dokumentasi

Dokumentasi berasal dari kata dokumen, yang berarti barang tertulis, metode dokumentasi berarti cara pengumpulan data dengan mencatat data-data yang sudah ada. Metode dokumentasi adalah mencari data mengenai hal-hal atau variabel yang berupa catatan buku, surat, transkrip, majalah, prasasti, notulen rapat, agenda dan sebagainya.

Teknik atau studi dokumentasi adalah cara pengumpulan data melalui peninggalan arsip-arsip dan termasuk juga buku-buku tentang pendapat, teori, dalil-dalil atau hukum-hukum dan lain-lain berhubungan dengan masalah penelitian.

d. Studi Literatur

Mempelajari tentang sistem informasi *retrieval* dan metode pencocokan string melalui berbagai macam media, antara lain melalui internet, jurnal-jurnal dan buku yang berhubungan dengan *text processing*.

e. Perancangan Sistem

Melakukan perancangan sistem dengan menguji algoritma yang digunakan terhadap data-data yang ada dan melakukan perhitungan manual apakah telah sesuai dengan yang diharapkan.

f. Implementasi

Pembuatan aplikasi pendeteksi plagiarisme berdasarkan perancangan

yang telah dibuat sebelumnya ke dalam program komputer.

- g. Uji coba produk dan evaluasi.
Melakukan uji coba program yang telah dibuat. Kemudian melakukan evaluasi terhadap kekurangan program dan memperbaikinya.

Jenis Data

Data yang dikumpulkan berupa data kuantitatif dan kualitatif. Sumber data penelitian ini adalah:

- a. Data primer merupakan data mentah yang diperoleh secara langsung melalui wawancara dengan pihak yang berhubungan dengan penelitian yang sedang dilakukan, kemudian akan diolah untuk tujuan tertentu sesuai dengan kebutuhan.
- b. Data sekunder adalah data-data yang didapat dari sumber bacaan dan berbagai macam sumber lainnya yang terdiri dari surat-surat pribadi, buku harian, not, sampai dokumen-dokumen resmi dari berbagai instansi pemerintah. Data sekunder juga dapat berupa majalah, buletin, publikasi dari berbagai organisasi, hasil-hasil studi, hasil survey, studi histories, dan sebagainya. Data sekunder dalam penelitian ini adalah berupa Skripsi Mahasiswa Fakultas Teknik, Jurusan Teknik Informatika Universitas Pamulang tahun 2012. Dan data yang digunakan untuk sampel adalah Bab I pada Skripsi Mahasiswa Fakultas Teknik, Jurusan Teknik Informatika.

Analisis Kebutuhan

Untuk mengetahui gambaran secara umum dari sistem yang akan dibangun, maka perlu dilakukan tahap analisis kebutuhan yaitu metode pengumpulan data, analisa kebutuhan perangkat lunak, dan analisa kebutuhan perangkat keras.

Analisis Kebutuhan Dokumen

Dokumen yang digunakan sebagai sampel adalah dokumen text Bahasa Indonesia yang berisikan karya ilmiah yang didapatkan dari hasil skripsi mahasiswa Fakultas Teknik, Jurusan Teknik Informatika Universitas Pamulang tahun 2012. Data yang digunakan untuk sampel adalah Bab I pada Skripsi Mahasiswa Fakultas Teknik, Jurusan Teknik Informatika. Pengembangan proses ini akan dilakukan memakai Algoritma Rabin-Karp dengan metode *Rolling Hash* untuk

menentukan tingkat akurasi pendeteksian plagiarisme. Sehingga dokumen text Bahasa Indonesia yang berisikan karya ilmiah ini nantinya akan digunakan untuk mengevaluasi tingkat keakuratannya. Dan diimplementasikan pada aplikasi pendeteksian plagiarisme.

Metode Pengumpulan Data

Teknik pengumpulan data pada penelitian ini menggunakan data sekunder yaitu data yang didapat tidak secara langsung dari objek penelitian. Data sekunder ini didapatkan pada proses dokumentasi seperti jurnal ilmiah, text books, dan internet yang berhubungan dengan penelitian yang dibahas oleh penulis.

Analisa Kebutuhan Perangkat Lunak

Perangkat Lunak yang digunakan antara lain:

- a. Operating Sistem (Windows XP SP2, Windows 7)
- b. XAMPP 1.6.3
- c. PHP Version 5.2.3
- d. Apache 2.0
- e. MySQL 5.0.45
- f. Heidi SQL 4.0
- g. PHP Designer 7.0
- h. Mozilla Firefox 3.6

Analisa Kebutuhan Perangkat Keras

Perangkat keras yang digunakan antara lain :

- a. Proceressor Core i3
- b. Ram 2 Gb
- c. Harddisk 500 Gb
- d. VGA NVIDIA 1 Gb
- e. Monitor 17"
- f. Motherboard dan keyboard

3. PEMBAHASAN

Pada tahap ini dibahas tentang perancangan sistem plagiarisme dengan menggunakan algoritma *Rabin-Karp* asli dan algoritma *Rabin-Karp* yang dimodifikasi. Adapun langkah-langkah yang dilakukan adalah:

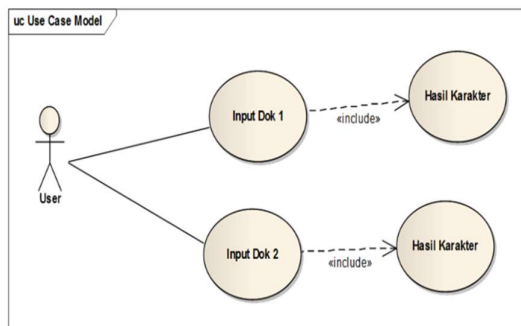
- a. Untuk membandingkan performa masing-masing algoritma maka dibuat aplikasi sederhana proses plagiarisme menggunakan algoritma *Rabin-Karp*.
- b. Menganalisa dan merancang sistem plagiarismedengan menggabungkan beberapa metode yang telah dijelaskan pada bab sebelumnya.
- c. Melakukan uji coba dengan 10 sampel karya ilmiah yang telah dibuat dengan

menganalisa hasil dari sistem algoritma tersebut. Hasil yang dikeluarkan oleh sistem berupa akurasi berupa presentasi dengan menggunakan algoritma *Rabin-Karp* yang asli dengan algoritma *Rabin-Karp* yang dimodifikasi.

Mengevaluasi hasil kedua algoritma tersebut apakah perbedaan yang dihasilkan dari algoritma *Rabin-Karp* yang asli dengan algoritma *Rabin-Karp* yang dimodifikasi. Pada tahap perancangan ini membahas mengenai pembuatan *Flowchart*.

Perancangan Use Case Diagram

Use Case diagram sistem pendeteksian plagiarisme :



Gambar 3.1 Use Case Diagram Sistem Pendeteksian Plagiarisme

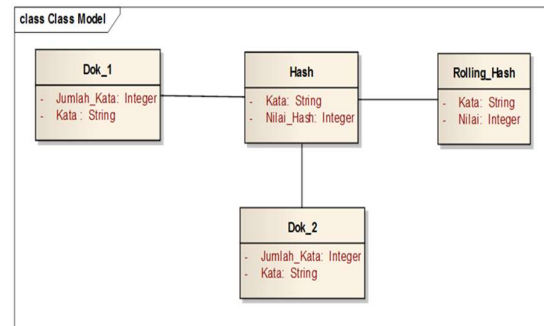
Use Case Diagram Sistem Pendeteksian Plagiarisme

Deskripsi Use Case:

- a Use Case : Input Keyword
- Actor : User
- Deskripsi :
 - User menginput dokumen << Include >>
 - User mencari dokumen 1 dan 2.
 - Setelah semua *field* diisi maka akan dilakukan proses pengecekan plagiarisme dokumen

Perancangan Class Diagram

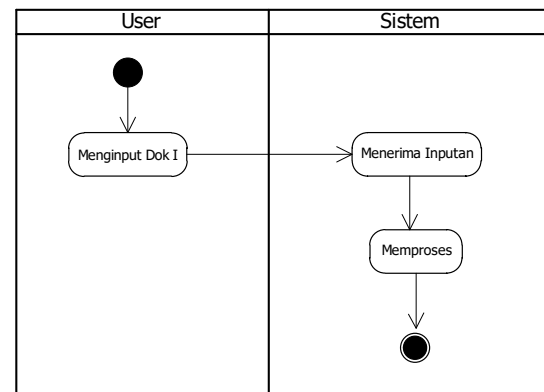
Berikut ini rancangan *Class Diagram* sistem pendeteksian plagiarisme :



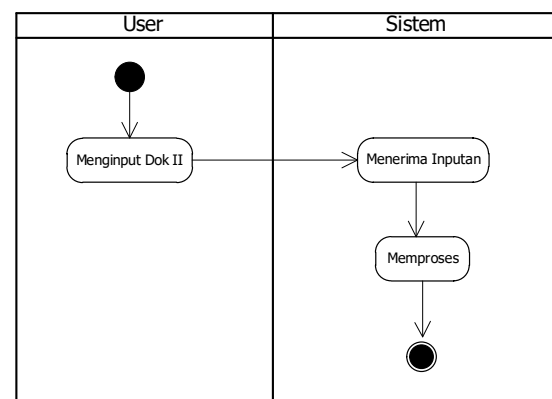
Gambar 3.2 Class Diagram Sistem Pendeteksian Plagiarisme

Activity Diagram

Berikut ini rancangan *activity* diagram sistem pendeteksian plagiarisme :



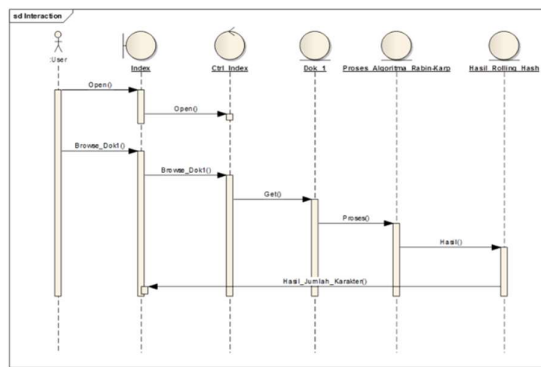
Gambar 3.3 Activity Diagram Input Dok I



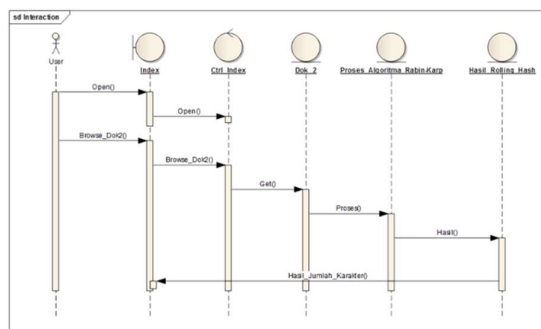
Gambar 3.4 Activity Diagram Input Dok I

Perancangan Sequence Diagram

Berikut ini rancangan *sequence* diagram sistem pendeteksian plagiarisme :



Gambar 3.5 Rancangan *Sequence Diagram* Dokumen 1



Gambar 3.6 Rancangan *Sequence Diagram* Dokumen 2

Perancangan User Interface

Pada sistem deteksi plagiarisme dokumen ini ada beberapa inputan dan parameter yang harus diisi oleh *user*, yaitu:

- Sumber dokumen I.
- Sumber dokumen yang akan diuji (II).
- Algoritma yang akan digunakan untuk proses deteksi plagiarisme dokumen.
- Nilai *rolling hash*.

Deteksi Plagiarisme
Algoritma Rabin-Karp

Upload File Text1 .doc :

Upload File Text2 .doc :

Gambar 3.7 Rancangan Layar Algoritma Rabin-Karp

Gambar 3.8 Rancangan Layar Hasil Algoritma Rabin-Karp

4. KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan yang disampaikan pada bab sebelumnya, maka terdapat beberapa kesimpulan sebagai berikut:

- Pendeteksian plagiarisme menggunakan algoritma *Rabin-Karp* dengan metode *rolling hash* dari hasil pengujian 30 dokumen teks yang sudah dijelaskan pada bab sebelumnya menghasilkan tingkat akurasi yang terbesar yaitu 47.58%. Hasil persentase tersebut termasuk dalam kategori tingkat plagiat 15-50%, berarti menandakan dokumen tersebut termasuk plagiat tingkat sedang. Sedangkan tingkat akurasi yang terkecil yaitu 19.28%, berarti menandakan dokumen tersebut termasuk plagiat tingkat sedang.
- Berdasarkan analisis proses pendeteksian tingkat plagiarisme menggunakan algoritma *raibin-karp* dengan metode *rolling hash* bisa membaca karakter berupa huruf, simbol seperti titik (.), koma (,), dan lain-lain.

5. SARAN

Ada beberapa hal yang perlu disampaikan agar menjadi bahan pertimbangan untuk melengkapi penelitian dimasa mendatang. Hal-hal tersebut diungkapkan dalam bentuk saran sebagai berikut:

- Apabila diperlukan, sistem pendeteksi plagiat ini dapat dikembangkan dengan penambahan metode baru untuk mencari solusi agar dapat menghasilkan nilai akurasi yang lebih akurat dalam pendeteksian plagiarisme.
- Aplikasi dapat dikembangkan agar tidak hanya menjadi *prototype* saja, melainkan diimplementasikan dalam proses pendeteksian secara nyata agar dapat

memberikan kemudahan dalam mendeteksi plagiarisme di suatu lembaga pendidikan ataupun di lembaga lainnya yang berhubungan dengan karya ilmiah.

DAFTAR PUSTAKA

- Arifin, Agus Zainal and Ari Novan Setiono. "Klasifikasi Dokumen Berita Kejadian Berbahasa Indoensia dengan Algoritma Single Pass Clustering." www.its.ac.id (2002).
- Arifudin, Nur. "Urgensi Penanganan Dokumen Perusahaan Melalui Sistem Yang Berteknologi Dilihat Dari Kekuatan Hukumnya." (2006).
- Asian, Jelita and Hugh E Williams. "Stemming Indonesia." Australia Computer Science Conference. 2005.
- ASKITIS, NIKOLAS and JUSTIN ZOBEL. "Redesigning the String Hash Table, Burst Trie, and BST to Exploit Cache." ACM Journal (2011).
- Baedlowi, Najib and Deka Aditia Adam. "String Matching dengan menggunakan algoritma Rabin Karp." Laboratorium Ilmu dan Rekayasa Komputasi, ITB (2006).
- Bernstein, Abraham and Christoph Kiefer. "Imprecise Queries Using Similarity Joins For Retrieval in Ontologis." iRDQL. Winterthurerstr: University Of Zurich, 2005.
- Cantone, Domenico and Simone Faro. "Fast-Search Algorithm New Efficient Variants Of The Boyer-Moore Pattern-Matching Algorithm." ACM Transaction (2003).
- Daptardar, Ajay and Dana Shapira. "Adapting the Knuth-Morris-Pratt Algorithm for Pattern Matching in Huffman Encoded Texts." ACM Transaction (2006).
- Fachrurrozi. Algoritma dan Pemrograman I. 2006.
- Firdaus, Hari Bagus. "Deteksi Plagiat Dokumen Menggunakan Algoritma Rabin-Karp." 2008.
- Garfield, Eugene. "A Tribute To Calvin N. Mooers, A Pioneer Of Information Retrieval." The Scientist (1997).
- Hasugian, Jonner. "Pengantar Kearsipan." n.d.
- Jimenez, Daniel A. "Neural Methods for Dynamic." Neural Methods for Dynamic (n.d.).
- Kalai, Adam. "Eficient Pattern-Matching with Don't Cares." Acm Transaction (2002).
- Mandala, Rila and Renaldi Munir. "Sistem Stemming Otomatis untuk kata dalam Bahasa Indonesia." Seminar Nasional Aplikasi Teknologi Informasi. Yogyakarta, 2004.
- Nugroho, Eko. "Perancangan Sistem Deteksi Plagiarisme Dokumen Teks Dengan Menggunakan Algoritma Rabin-Karp." 2011.
- Nugroho, Yohanes. "Hashtable dan vektor bit." 2002.
- Petrini, Fabrizio, Virat Agarwal and Davide Pasetto. "SCAMPI: a Scalable CAM-based Algorithm for Multiple Pattern Inspection." ACM Journal (2009).
- Plaxton, Greg. "String Matching: Rabin-Karp Algorithm." Theory in Programming Practice, University of Texas at Austin 2005.
- Putra, Dian Perdhana. "Perancangan Algoritma Pencocokan String Menggunakan Fungsi Hash Untuk Pendeteksian Plagiarisme." 2010.
- Rance D., Necaise. "Data Structures and Algorithms Using Python." JOHN WILEY & SONS, INC. Data Structures and Algorithms Using Python. England: Department of Computer Science College of William and Mary, 2011.
- Renier, G.J. History its Purpose and Method. Yogyakarta: (terjemahan Muin Umar). Yogyakarta: Pustaka Pelajar, 1997.
- Shalev, Ori and Nir Shavit. "Ssplit Order List Extensible Hash Tables." ACM Transaction (2006).
- Singla, Nimisha and Deepak Garg. "String Matching Algorithms and their Applicability in various Applications." International Journal of Soft Computing and Engineering (IJSCE) (2012).
- Subekti, Nanang Bagus and Asep Ahmad Fauzi. "Maraknya Aksi Plagiarisme Di Indonesia." 3 Maret 2012.