

SISTEM DETEKSI KEMIRIPAN ANTAR DOKUMEN TEKS MENGUNAKAN ALGORITMA SCAM

Hevi Oktaviani, 12 1065 1056
Program Studi Teknik Informatika

Abstrak

SCAM adalah singkatan dari *Stanford Copy Analysis Mechanism* dimana menjadi biasanya tolak ukur relatif untuk mendeteksi *overlapping* dengan membuat perbandingan pada satu set kata-kata yang umum antara dokumen tes dan dokumen terdaftar.

Vector Space Model sebuah model populer dalam domain IR, adalah model VSNI. Mengingat permintaan dengan bobot yang sesuai, produk dot terjadinya vektor tertimbang dari query dengan dokumen yang disimpan dan dihitung: jika nilai produk dot melebihi batas tertentu, dokumen ditandai untuk mencocokkan query.

Sistem ini mampu melakukan pengukuran kemiripan dokumen Bahasa Indonesia. Dokumen yang digunakan berupa file import dokumen dan input text jadi user dapat memilih input data dengan.

Kata Kunci : Scam, *overlapping*, *Vector Space Model*.

1. Latar Belakang

Digitalisasi pengolahan informasi dengan menggunakan komputer menghasilkan fasilitas yang *copy-paste* (salin-tempel) sehingga memudahkan pengolahan data sesuai dengan kebutuhan misalnya memenuhi tugas kuliah, membuat paper, dan sebagainya. Hal ini tentu berpotensi terjadinya tindakan penjiplakan suatu karya tulis tanpa ijin seperti plagiat. Selain itu, semakin meningkat kecanggihan manusia dalam melakukan plagiat juga menjadi alasan lain dibutuhkannya sistem yang dapat membantu mendeteksi plagiat dalam sebuah dokumen. Oleh karena itu, penulis menggagas untuk dibangunnya sebuah sistem deteksi plagiat (*Plagiarism detection system*) yang menggunakan SCAM (*Stanford Copy Analysis Mechanism*) untuk melakukan pencocokan kemiripan antar dokumen dengan menyajikan dokumen plagiat yang disusun berdasarkan kemiripan dokumen. SCAM sendiri mempunyai ukuran relatif yang dapat mendeteksi tumpang tindih (*overlap*). Dengan adanya sistem deteksi plagiat ini diharapkan akan membantu proses deteksi plagiat.

2. Penerapan Pengukuran Kesamaan Dokumen

Pengukuran kemiripan dokumen dapat digunakan untuk mengukur kemiripan dokumen resmi, dokumen standar seperti

Standar Operasional Procedure (SOP), peraturan perundangan, hasil penelitian dan lain-lain. Pengukuran dokumen dapat diterapkan di berbagai instansi seperti pemerintahan, perusahaan, akademik dan lain-lain. Plagiarisme dalam bidang akademik dapat dibagi menjadi dua (Vamplew & Dermaoudly, 2015) yaitu :

- a. *Content-based file comparison*
Pendekatan *content-based file* merupakan pendekatan untuk mengukur kemiripan dokumen atau teks seperti tugas *essay* mahasiswa dan hasil penelitian.
- b. *Content-based comparison of source code*
Pendekatan ini digunakan untuk mengukur kemiripan *source code* pemrograman.

3. Pengukuran Kemiripan

Pada sub bab ini dijelaskan tentang jenis pengukuran kemiripan, pendekatan pengukuran kemiripan, penelitian-penelitian yang telah dilakukan terhadap pendekatan pengukuran kemiripan.

4. Algoritma SCAM

SCAM adalah singkatan dari *Stanford Copy Analysis Mechanism* dimana menjadi biasanya tolak ukur relatif untuk mendeteksi *overlapping* dengan membuat perbandingan pada satu set kata-kata yang

umum antara dokumen tes dan dokumen terdaftar.

Secara general algoritma ini bekerja dengan memberikan pendekatan-pendekatan khusus untuk mendeteksi kesamaan antara dokumen. Dan langkah-langkah deteksi tersebut, dibagi menjadi empat langkah utama, yaitu:

1. Pengindeksan *dataset* : Dataset ini terdiri dari beberapa dokumen yang diproses dan di indeks.
2. Pengolahan dokumen uji : Dokumen yang akan diuji diberikan dalam input diproses untuk proses *tokenize* yang akan dimasukkan dalam daftar kata yang akan diolah nantinya.
3. Pencarian pada indeks : Indeks akan digunakan untuk mengambil kecocokan antara dokumen uji dan dokumen milik *dataset*.
4. Mengevaluasi kesamaan: Menggunakan rumus SCAM. perbandingan kesamaan dokumen, nantinya akan berdasarkan dokumen uji dan dokumen milik *dataset*.

5. Implementasi Sistem

Berdasarkan hasil analisis dan perancangan sistem yang telah dilakukan, maka dilakukan implementasi sistem deteksi kemiripan dokumen dengan metode SCAM ke dalam bentuk program dengan menggunakan bahasa pemrograman PHP dan database MySQL.

5.1 Tampilan Utama Sistem

Tampilan utama merupakan halaman yang akan dijalankan ketika sistem dijalankan, halaman utama menampilkan nama aplikasi dan penyusun penelitian, halaman utama dapat dilihat seperti gambar dibawah ini:



Gambar 4.1. Halaman Utama

5.2 Halaman Data Set

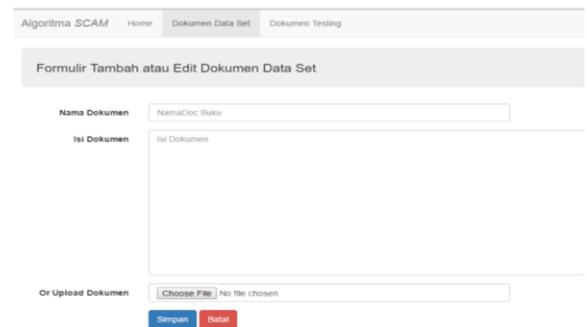
Pada tampilan menu Data Set, akan menampilkan sebuah tabel dimana Pengguna dapat memilih berapa jumlah dataset yang digunakan, pada halaman ini user disediakan beberapa fitur yaitu Tambah, Edit dan Hapus. Tampilan

halaman dataset dapat kita lihat pada gambar 4.2.

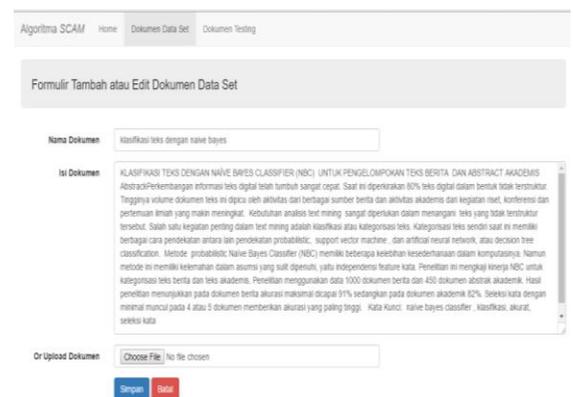


Gambar 4.2. Halaman Dataset

Pada gambar 4.2 ada beberapa fitur yaitu cari dokumen digunakan untuk mencari dokumen dataset, tombol tambah digunakan untuk membuka halaman form tambah dokumen dataset seperti gambar 4.3, tombol edit merupakan fitur untuk melakukan perubahan dataset dan jika user memilih hapus maka user akan menghapus dataset.



Gambar 4.3. Halaman Form Tambah Dokumen Dataset



Gambar 4.4 Halaman form edit data set

5.3 Halaman Testing

Menu halaman testing akan menampilkan sebuah form input berupa teks dokumen uji, dimana sistem akan menggunakan input text atau import dokumen untuk melakukan testing kesamaan dokumen sistem akan mengolah data yang

