

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Plagiarisme atau penjiplakan hasil karya seseorang dan mengakuinya sebagai hasil karya pelaku plagiarisme (plagiator) merupakan suatu tindakan yang melanggar hukum, hal ini tercantum pada Undang-Undang Republik Indonesia, tepatnya pada UU No. 20 Tahun 2003. Bagi pelanggar yang terbukti melakukan tindakan plagiarisme pada karya ilmiahnya, maka akan dilakukan pencabutan gelar seperti yang tercantum pada Pasal 25 Ayat 2. Selain pencabutan gelar, pelaku plagiarisme juga akan dikenakan sanksi pidana penjara selama dua tahun dan/atau denda paling banyak dua ratus juta rupiah seperti yang tercantum pada Pasal 70 [1].

Plagiarisme pada kode sumber, bisa saja terjadi pada program studi S1 Teknik Informatika. Untuk melakukan penilaian keaslian kode sumber, diperlukan pendekatan *attribute-based-approach* ataupun *structure-based-approach* [2]. Prinsip kerja dari pendekatan *attribute based* ialah mengubah kode sumber dalam bentuk *intermediate representation*, seperti *token* kode sumber setelah itu membandingkan kemunculan *attribute* pada pasangan kode sumber dan menilai hitung similaritasnya. Sedangkan prinsip kerja dari *structure based* ialah mengubah kode sumber dalam bentuk *intermediate representation* seperti *token* kode sumber setelah itu akan dibandingkan menggunakan algoritma *string-matching*. Terdapat beberapa algoritma yang cukup populer dan diciptakan untuk membandingkan kesamaan kode sumber. Algoritma tersebut beberapa di antaranya ialah algoritma *Running-Karp-Rabin Greedy-String-Tiling* (RKRGST) [3] dan algoritma *Local Alignment* [4]. Meskipun algoritma RKRGST merupakan algoritma yang cukup populer untuk mendapatkan nilai kesamaan dari pasangan kode sumber hasil plagiarisme, akan tetapi algoritma ini memiliki waktu proses yang cukup lama. Hal ini dikarenakan algoritma RKRGST memiliki kompleksitas $O(n^3)$ [3]. Selain dari waktu proses yang cukup lama, nilai similaritas yang dihasilkan algoritma ini dirasa kurang sensitif terhadap jumlah *subsequences* yang terbentuk. Hasil nilai similaritas dari satu *sequence* panjang dan kumpulan *sequence* kecil dapat bernilai sama walaupun *sequence* panjang seharusnya lebih cenderung dianggap hasil

plagiarisme. Hal ini akan menyebabkan pengajar atau asisten dosen kesulitan untuk menentukan pasangan kode sumber mana yang melakukan plagiarisme.

Berdasarkan paparan masalah yang diungkapkan, maka dibutuhkan penerapan penggabungan *attribute-based approach* menggunakan *Vector Space Model* (VSM) sebagai tahap *filtering initial pair* di mana jika nilai similaritas VSM yang dihasilkan melebihi dari batas akan dihitung nilai similaritasnya menggunakan algoritma RKGST. *Filtering initial pair* ini juga akan diterapkan pada algoritma *Local Alignment*. Sedangkan untuk menghasilkan nilai similaritas yang sensitif terhadap jumlah *sequence*, maka dibutuhkan persamaan normalisasi baru dari algoritma RKGST yang sensitif akan jumlah *subsequences*. Persamaan normalisasi yang dihasilkan akan diujicobakan terhadap algoritma RKGST. Normalisasi yang dihasilkan akan menerapkan konsep penalti seperti yang dimiliki oleh algoritma *Local Alignment* dan akan diujicobakan pada perhitungan nilai similaritas dengan metode *average* dan *maximum* dari algoritma RKGST. Persamaan ini akan menghasilkan nilai similaritas yang lebih akurat pada kasus perubahan urutan proses. Selain itu, untuk mempresentasikan hasil perhitungan dari algoritma yang ada, maka dibangun pula aplikasi *prototype* serta akan dilakukan pula uji *dataset* terkontrol dan empirik untuk mempelajari karakteristik dari persamaan normalisasi.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang diulas, dapat ditarik rumusan masalah sebagai berikut:

1. Bagaimana mempercepat waktu proses komparasi perhitungan similaritas kode sumber?
2. Bagaimana merancang *normalized similarity equation* yang sensitif akan *subsequences*?

1.3 Tujuan Pembahasan

Adapun tujuan dari penelitian ini adalah:

1. Menerapkan penggabungan *attribute-based approach* sebagai *initial filtering* dengan menerapkan pemodelan *Vector Space Model*.

2. Merancang persamaan normalisasi yang sensitif akan jumlah *subsequences* dengan memberikan penalti yang berbanding lurus dengan jumlah *subsequences* terbentuk.

1.4 Ruang Lingkup

Ruang lingkup untuk tugas akhir ini, meliputi beberapa hal:

1. Hasil dari perhitungan algoritma hanya menunjukkan kesamaan kode program tanpa menentukan apakah kode program tersebut merupakan hasil plagiarisme atau tidak.
2. Aplikasi dapat mendeteksi kesamaan dari dua kumpulan kode program selama tertulis dalam bahasa pemrograman yang sama (Java/Python/C++).
3. Karakteristik hanya didasarkan dari data evaluasi sehingga masih memungkinkan adanya karakteristik yang tidak dibahas.

1.5 Sumber Data

Sumber data yang diperoleh dalam pembuatan tugas akhir ini adalah studi literatur terhadap penelitian sebelumnya yang menggunakan algoritma *Running-Karp-Rabin Greedy-String-Tiling* dan *Local Alignment*, sistem deteksi plagiarisme kode sumber yang sudah ada, *dataset* dari publikasi ilmiah terkait.

1.6 Sistematika Penyajian

Sistematika penulisan yang akan digunakan pada laporan ini adalah sebagai berikut:

BAB I PENDAHULUAN

Bab ini menjelaskan tentang latar belakang, rumusan masalah, tujuan pembahasan, ruang lingkup kajian, sumber data dan sistematika penulisan yang berhubungan dengan judul tugas akhir.

BAB II LANDASAN TEORI

Bab ini berisi seluruh teori yang dipakai dan relevan dengan pengembangan penelitian yang dibahas dalam laporan ini.

BAB III ANALISIS

Bab ini berisikan analisis dari fitur-fitur yang akan tersedia pada program pendeteksian plagiarisme kode sumber.

BAB IV IMPLEMENTASI

Bab ini berisi *screenshot* aplikasi dan implementasi pemodelan yang telah dibuat pada bab sebelumnya.

BAB V PENGUJIAN

Bab ini berisi pengujian dan evaluasi dari program pendeteksian plagiarisme kode sumber.

BAB VI PENUTUP

Bab ini terdiri dari simpulan mengenai penelitian yang telah dibuat dan saran untuk perbaikan serta penyempurnaan penelitian ini.

