

BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1. Tinjauan Pustaka

Beberapa penelitian sebelumnya yang telah dilakukan terkait aplikasi antrian dan implementasi qr code, antara lain sebagai berikut :

Menurut penelitian yang dilakukan Agustinus Philipus Tri Yunianta Arum Surya (Agustinus 2017), pembuatan sistem deteksi kemiripan kata pada dua dokumen menggunakan Algoritma Rabin-Karp berhasil dibangun dengan berbasis web. Sistem ini dapat menghasilkan persentase kemiripan kata dan mengindikasikan kalimat sama pada dua dokumen. Hasil perhitungan tingkat keakuratan sistem memiliki nilai keakuratan 95% diperoleh dari 20 dokumen pengujian identifikasi kalimat sama dengan rata-rata waktu proses eksekusi sistem sebesar 11,6 detik dari 20 dokumen pengujian sistem deteksi kemiripan kata. Semakin banyak kata yang diproses, maka akan semakin lama proses eksekusi. Namun dalam penelitian Agustinus Philipus Tri Yunianta Arum Surya, belum menyediakan sistem yang dapat mendeteksi banyak dokumen.

Menurut penelitian yang dilakukan Panji Novantara 1 , Opim Pasruli 2 (Panji dan Opim 2018), Sistem pendeteksi plagiarisme dengan algoritma Jaro-Winkler Distance, waktu proses yang dibutuhkan dalam melakukan pendeteksian plagiarisme bergantung pada banyaknya isi dokumen, dan ukuran file yang akan diproses.

Menurut penelitian yang dilakukan Dedi Leman 1 , Gunadi Widi Nurcahyo

2 , Sarjon Defit 3 (Dedi, Gunadi dan Sarjon 2015), Aplikasi pendeteksian plagiarisme ini dirancang menggunakan bahasa pemrograman java dan IDE Netbeans 6.0.1. Pada uji coba ini terdapat 3 dokumen uji yang mempunyai 3 dokumen latih dengan algoritma Rabin-Karp dan serta menggunakan *stemming*. Hasil pengujian akurasi similaritas atau tingkat kesamaan dengan algoritma. Pada kode dokumen A dengan dokumen latih 100% sama menghasilkan similarity 100% sama dengan kecepatan waktu atau streaming 0.56 T(s). Kemudian pada dokumen latih 90% sama menghasilkan similarity 90.98% dengan kecepatan waktu 0.45 T(2). Sedangkan dokumen latih 40% sama menghasilkan similarity 40.7% dengan kecepatan waktu 0.53 T(s). Penggunaan *stemming* berpengaruh pada akurasi nilai similarity yang dihasilkan. Dengan menggunakan *stemming* menghasilkan nilai yang cenderung kurang baik dibandingkan tanpa menggunakan *stemming*.

Sedangkan menurut yang dilakukan (Sunardi 1, Anton Yudhana 2, dan Iif Alfiatul Mukaromah 3, 2018). Hasil akhir dari pendeteksi plagiarisme pada gambar 8 mampu menghasilkan presentasi sampai 100% dengan menggunakan *Jaccard Similarity* dari dokumen atau sample pada Gambar 6 dengan nilai *n-gram* 3 dan nilai *w-gram* 5. Jika dibandingkan dengan metode *k-gram* yang digunakan pada [16] yang memperoleh kemiripan 83%, hasil ini cukup andal. Dengan demikian, metode *Jaccard Similarity* memiliki prospek untuk digunakan dalam deteksi plagiat.

Usulan (Azuwar Edison 2020). Hasil akhir dari deteksi plagiarisme ini mampu mendeteksi menghasilkan persentase sesuai dengan yang dibahas pada bab sebelumnya yaitu dengan melakukan pengujian kemiripan judul skripsi 100%

mirip, 50% mirip dan 30% mirip berhasil dilakukan. Dalam penelitian ini didapatkan hasil kesimpulan perbedaan dengan penelitian sebelumnya yakni hasil proses yang didapatkan hanya *N-Grams*, *Rolling Hash*, *Window*, *Fingerprints* dan *Similarity*. Dari hasil pengujian yang dilakukan, pengguna tidak dapat mengetahui berapa lama waktu proses yang dibutuhkan untuk mendeteksi kemiripan judul skripsi (*StopWord*). Dari hasil kemiripan 100% mirip akan menghasilkan persentase berbeda jika data judul skripsi 100% mirip dibolak-balikkan. Hasil yang didapat adalah diatas 80%.

Tabel 2. 1 Tinjauan Pustaka

No.	Sumber	Objek	Metode Penelitian	Hasil
1.	Agustinus Pilipus Triyunianta Arum Surya (2017)	Deteksi Plagiarisme Karya Tulis Ilmiah	Algoritma Rabin-Karp	Akurasi 95% menggunakan 20 dokumen waktu 11 detik
2.	Panji Novantara* 1 , Opin Pasruli 2 (2018)	Dokumen Skripsi	Algoritma Jaro-Winkler Distance	Waktu proses yang dibutuhkan dalam melakukan pendeteksian plagiarisme bergantung pada banyaknya isi dokumen, dan ukuran file yang akan di proses.
3.	Dedi Leman 1 , Gunadi Widi Nurcahyo 2 , Sarjon Defit 3 (2015)	Dokumen Teks	Algoritma Rabin-Karp	Dokumen A dengan dokumen latih 100% sama, hasil similarity 100% sama dengan waktu 0.56 T(s) Dokumen latih 90% sama, dengan hasil

No.	Sumber	Objek	Metode Penelitian	Hasil
				similarity 90.98% dengan kecepatan waktu 0.45 T(s) Dokumen latih 40% sama, dengan hasil similarity 40.7% dengan kecepatan waktu 0.53 T(s)
4.	Sunardi 1, Anton Yudhana ² , dan Iif Alfiatul Mukaromah ^{3*})	File Dokumen	Algoritma WInnowing	Hasil akhir dari pendeteksi plagiarisme menghasilkan presentasi sampai 100% menggunakan Jaccard Coefficient dengan n-gram 3 dan nilai w-gram 5
5.	Afni Rul Jannah. R 2018	Judul Skripsi	Algoritma String Matching	Sistem deteksi kemiripan judul berhasil menghasilkan output persentase kemiripan berdasarkan judul yang dimasukkan
6.	Usulan 2020	Judul Skripsi	<i>Algoritma WInnowing dan Jaccard Coefficient</i>	Dari pengujian yang dilakukan menggunakan <i>Algoritma WInnowing</i> dapat diketahui hasil dari <i>N-Grams, Rolling, Window</i> dan <i>Fingerprints</i> .

2.2. Dasar Teori

2.2.1. WInnowing

WInnowing adalah algoritma yang digunakan untuk melakukan proses pengecekan kesamaan kata (document fingerprinting) untuk mengidentifikasi

plagiarisme (penjiplakan). Document fingerprinting merupakan metode yang digunakan untuk mendeteksi keakuratan salinan antar dokumen atau hanya sebagian teks saja. Prinsip kerja dari metode document fingerprinting ini adalah dengan menggunakan teknik hashing. Teknik hashing adalah sebuah fungsi yang mengkonversi setiap string menjadi bilangan. Adapun langkah-langkah algoritma winnowing, sebagai berikut (Afni Rul Jannah. R 2018) :

1. Penghilangan karakter umum seperti tanda baca, kata sambung, simbol matematik, spasi antar kata, serta menyeragamkan huruf menjadi lowercase, seperti contoh dibawah ini :

```
Hallo !!! Saya Azuwar
```

Akan diubah menjadi :

```
hallosayaazuwar
```

2. Pembentukan N-Gram dengan nilai gram yang telah ditentukan sehingga menghasilkan rangkaian gram. seperti contoh dibawah ini dengan N-gram 5 :

```
hallo allos llosa losay osaya sayaa ayaaaz
yaazu aazuw azuwa zuwar
```

3. Pengkodean tiap gram dengan menggunakan nilai ASCII dari masing-masing karakter pada suatu gram, proses ini disebut dengan Rolling hash.

$$H(ck) = c_1 * b^{(k-1)} + c_2 * b^{(k-2)} + \dots + c_k * b^{(k-k)}$$

Keterangan :

c = nilai ascii karakter

b = basis (bilangan prima)

k = banyak karakter

hasil rolling hash dari kalimat diatas adalah :

6398	6370	6726	6782	6846	6782	6448
6922	6338	6662	7344			

4. Pembentukan rangkaian Window dari rangkaian gram yang telah dikodekan yang kemudian akan dicari nilai fingerprint dari tiap-tiap rangkaian Window tersebut. Pembentukan nilai *hash* dari window dengan ukuran $w = 4$ yaitu

W-1	:	{	6398	6370	6726	6782	}
W-2	:	{	6370	6726	6782	6846	}
W-3	:	{	6726	6782	6846	6782	}
W-4	:	{	6782	6846	6782	6448	}
W-5	:	{	6846	6782	6448	6922	}
W-6	:	{	6782	6448	6922	6338	}
W-7	:	{	6448	6922	6338	6662	}
W-8	:	{	6922	6338	6662	7344	}

5. Pengukuran nilai Similarity, pada proses ini akan menghasilkan suatu nilai irisan dari populasi fingerprint antara judul skripsi yang dicari dengan sumber judul skripsi pembanding.

6370	6370	6726	6448	6448	6338	6338	6338
------	------	------	------	------	------	------	------

6. Nilai *fingerprint* yang dibentuk dari algoritma Winnowing digunakan untuk mengukur prosentase kemiripan teks pada persamaan *Jaccard Coefficient*. Berikut rumusnya :

$$\text{Similarity} = (\text{Jumlah_Fingerprint_sama} / \text{Total_seluruh_fingerprint}) \times 100$$

2.2.2. Jaccard Coefficient

Jaccard Coefficient adalah salah satu metode yang dipakai untuk menghitung similarity antara dua objek. Ditemukan oleh Paul Jaccard yang merupakan metode ukuran kesamaan yang digunakan untuk membandingkan kesamaan dan keragaman 2 set sampel. Hasil dari perhitungan akan dihasilkan nilai similaritas. Nilai similaritas yang tertinggi dapat dianggap bahwa nilai tersebut paling similar, dengan kata lain memiliki banyak kesamaan. Jaccard Coefficient menghitung similarity antara dua objek, A dan B yang dinyatakan dalam dua buah vektor.

$$\text{Jaccard Coefficient (A,B)} = \frac{|(A \cap B)|}{|(A \cup B)|} \times 100$$

Keterangan :

A = Hash dokumen 1 (A₁, A₂, A₃,..... A_n)

B = Hash dokumen 2 (B₁, B₂, B₃,..... B_n)

Dimana (A,B) adalah nilai similarity, $|A \cap B|$ adalah jumlah dari hash yang sama dari dokumen 1 dan 2 dan $|A \cup B|$ adalah jumlah hash dari dokumen 1 dan dokumen 2. Sebagai contoh A = {1,2,4}, b = {1,2,4,7,8} maka, $|A \cap B| = \{1,2,4\}$ dan $|A \cup B| = \{1,2,4,7,8\}$ sehingga :

$$(A,B) = \frac{3}{5} \times 100 = 60$$

Dalam kasus menghitung persentase kemiripan rumusnya disesuaikan menjadi seperti pada persamaan :

Jaccard Coefficient (A,B) =

$$\frac{\text{Jumlah kata yang sama}}{\text{Jumlah kata yang berbeda}} \times 100$$

2.2.3. Similarity

Similarity merupakan tingkat perbandingan persentase kemiripan antar dokumen yang diuji. Similarity ini akan dijadikan acuan dalam menentukan persentase tingkat kemiripan sebuah dokumen yang diuji. Besarnya persentase similarity akan dipengaruhi oleh tingkat kemiripan dari dokumen yang diuji, semakin besar persentase similarity maka tingkat kemiripan akan semakin dianggap tinggi. Konsep similarity sudah menjadi isu yang sangat penting di hampir setiap bidang ilmu pengetahuan. Menurut Zaka, (2009) dalam disertasinya menjelaskan tiga macam teknik yang dibangun untuk menentukan nilai similarity (kemiripan) dokumen. Similarity atau kemiripan dalam kata adalah mencari kesamaan antara kata inputan dengan kata sumber. Kemiripan yang dimaksud bukanlah sinonim, melainkan kata yang mendekati kata sumber (kata sebenarnya).

Kesamaan teks harus menentukan seberapa 'dekat' dua bagian teks dalam kedekatan permukaan (kesamaan leksikal) dan makna (kesamaan semantik). Misalnya, seberapa miripkah frasa “guru memberikan pidatonya di sebuah ruang kosong” dengan “hampir tidak orang ketika guru berbicara”. Meskipun menyampaikan arti yang sangat mirip, ini ditulis dengan cara yang sangat berbeda.

Faktanya, dari kedua kalimat tersebut hanya memiliki satu kata yang sama (“guru”), dan bukan yang benar-benar signifikan.

2.2.4. Plagiarisme

Plagiarisme atau sering disebut plagiat adalah penjiplakan atau pengambilan karangan, pendapat, dan sebagainya dari orang lain dan menjadikannya seolah karangan dan pendapat sendiri. Plagiat dapat dianggap sebagai tindak pidana karena mencuri hak cipta orang lain. Di dunia pendidikan, pelaku plagiarisme dapat mendapat hukuman berat seperti dikeluarkan dari sekolah/universitas. Pelaku plagiat disebut sebagai plagiator. Singkat kata, plagiat adalah pencurian karangan milik orang lain. Dapat juga diartikan sebagai pengambilan karangan (pendapat dan sebagainya) orang lain yang kemudian dijadikan seolah-olah miliknya sendiri. Setiap karangan yang asli dianggap sebagai hak milik si pengarang dan tidak boleh dicetak ulang tanpa izin yang mempunyai hak atau penerbit karangan tersebut. Sesudah 2 × 24 jam berita surat kabar tersiar, maka seseorang dapat mengambil alih dengan syarat harus menyebutkan sumbernya.

Plagiarisme merupakan salah satu “kejahatan intelektual” yang terjadi di dalam dunia akademik, kejahatan tersebut dapat tergambar dari perilaku pencurian, penipuan, penulisan dan pengakuan hasil penelitian maupun tulisan orang lain yang kemudian diakui sebagai karya sendiri. Banyak kalangan mahasiswa yang melakukan plagiarisme dalam mengerjakan tugas-tugasnya, hal ini tentunya sangat tidak baik, mahasiswa tidak mau berfikir sendiri untuk menyelesaikan tugasnya. Untuk meminimalisir terjadinya tingkat plagiarisme di kalangan mahasiswa maka dibuatlah aplikasi yang bisa untuk mendeteksi plagiarisme yang dilakukan oleh

mahasiswa, aplikasi berbasis website ini memanfaatkan algoritma winnowing, yaitu salah satu metode document fingerprinting yang digunakan untuk mendeteksi kemiripan antara teks dokumen dengan menggunakan teknik hashing (Susanto, E. H., & Aslamiyah, S. 2017). Plagiarisme juga tidak mengacu kepada hasil karya tulisan saja melainkan juga hasil karya musik, desain, dll.

Berikut definisi dan pengertian plagiarisme dari beberapa sumber buku :

1. Menurut Lindsey, plagiat adalah tindakan menjiplak ide, gagasan atau karya orang lain untuk diakui sebagai karya sendiri atau menggunakan karya orang lain tanpa menyebutkan sumbernya sehingga menimbulkan asumsi yang salah atau keliru mengenai asal muasal dari suatu ide, gagasan atau karya (Soelistyo, 2011).
2. Menurut Suyanto dan Jihad (2011), plagiarisme adalah mencuri gagasan, kata-kata, kalimat, atau hasil penelitian orang lain dan menyajikannya seolah-olah sebagai karya sendiri.
3. Menurut Brotowidjoyo (1993), plagiarisme merupakan pembajakan berupa fakta, penjelasan ungkapan dan kalimat orang lain secara tidak sah.
4. Menurut Ridhatillah (2003), plagiarisme adalah tindakan penyalahgunaan, pencurian atau perampasan, penerbitan, pernyataan atau menyatakan sebagai milik sendiri sebuah pikiran, ide, tulisan, atau ciptaan yang sebenarnya milik orang lain.