

BAB III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilakukan di kampus Politeknik Negeri Malang. Penelitian dilaksanakan selama 7 bulan dimulai pada bulan Desember 2020 sampai dengan Juni 2021.

3.2 Metode Pengumpulan Data

Data penelitian yang diolah pada penelitian ini adalah dokumen teks berupa jurnal. Data tersebut akan diproses untuk mengetahui apakah dokumen teks tersebut memiliki kemiripan satu sama lain. Metode pengumpulan data yang digunakan dalam pembuatan sistem ini adalah:

A. Studi Literatur

Pengumpulan data dengan melakukan studi pustaka mencakup artikel, literatur dan buku yang berhubungan dengan merancang sebuah sistem pendeteksi plagiarisme dan metode *Winnowing*. Artikel jurnal yang dijadikan acuan dalam proses pengolahan data menggunakan metode *Winnowing* adalah jurnal yang diteliti oleh Nurdin dan Amin Munthola yang berjudul Sistem Pendeteksian Kemiripan Judul Skripsi Menggunakan Algoritma *Winnowing*. Pada jurnal tersebut menjelaskan secara detail langkah metode *Winnowing* beserta contoh pengolahannya.

B. Observasi

Metode observasi atau pengamatan pada penelitian ini dilakukan dengan mencari dan mengamati jurnal yang sudah ada dari situs seperti *ieee explore*, *sciencedirect* dan *google scholar* yang akan digunakan untuk menguji tingkat plagiarisme. Data yang dibutuhkan adalah jurnal uji dan jurnal pembanding dengan format PDF maupun teks. Data yang digunakan dalam pengujian dituliskan pada Tabel 5.2 di mana diambil dari beberapa sumber seperti ieeexplore.ieee.org, jurnal.wicida.ac.id, stmikplk.ac.id, citeseerx.ist.psu.edu, aclweb.org, jip.polinema.ac.id, ejournal.uigm.ac.id, journal.uniku.ac.id, journal.unj.ac.id, sisfotenika.stmikpontianak.ac.id, ojs2.unwahas.ac.id, ejournal.gunadarma.ac.id, jurnalnasional.ump.ac.id, researchgate.net, semanticscholar.org, link.springer.com, journal.upgris.ac.id, journal.ubm.ac.id, jurnal.umk.ac.id, cursorjournal.org,

jurnal.informatika.petra.ac.id, publikasi.dinus.ac.id, research.pps.dinus.ac.id, journal.unilak.ac.id, journal.untar.ac.id, njca.co.id dan knsi.stikom-bali.ac.id.

3.3 Metode Pengolahan Data

Metode *Winnowing* dapat digunakan dalam mendeteksi plagiarisme dokumen teks lintas bahasa. Hasil dari pengolahan data menggunakan metode ini berupa nilai persentasi kemiripan. Berikut merupakan langkah-langkah dalam perhitungan metode *Winnowing*:

3.3.1 Input Dokumen

Dokumen yang dapat dimasukkan dalam *database* dapat berupa teks dan *file* PDF berbahasa Inggris dan Indonesia. Dokumen uji merupakan dokumen yang dijadikan acuan dalam pengujian, sedangkan dokumen pembanding merupakan dokumen yang dijadikan perbandingan dalam pengujian untuk dicari persentase kemiripan antar dokumen tersebut.

3.3.2 Menerjemahkan Dokumen Teks

Jika dokumen yang diproses berbahasa Inggris maka dokumen tersebut akan diterjemahkan terlebih dahulu ke dalam bahasa Indonesia. Dokumen yang dapat dideteksi hanya dokumen berbahasa Indonesia dan berbahasa Inggris.

3.3.3 Pembatasan Uji Dokumen

Setelah proses terjemahan, langkah selanjutnya adalah dengan melakukan pembatasan pada dokumen yang diuji yaitu dengan tidak mengikutkan kalimat yang berada dalam tanda petik, sitasi dan referensi. Hal ini bertujuan untuk memaksimalkan dalam proses deteksi kesamaan teks.

3.3.4 *Preprocessing*

Setelah dokumen berhasil diterjemahkan, maka langkah selanjutnya yaitu melakukan tahap *preprocessing*. Pada *preprocessing* terdapat 3 tahap yang dilakukan yaitu *case folding*, *tokenizing* dan *filtering*. Teks yang di masukkan akan diproses menjadi huruf kecil dan menghilangkan karakter atau tanda baca yang tidak relevan sehingga hanya menampilkan huruf saja dan membuang kata yang tidak penting.

3.3.5 Membentuk Nilai *N-Gram*

Hasil dari preprocessing akan menghasilkan data murni yang akan digunakan pada proses *n-gram* membentuk *substring*. Tahap ini merupakan pembentukan rangkaian karakter sepanjang *n* dan akan digunakan untuk menghitung *hash*. Nilai *n* yang digunakan adalah 9 dan dalam pemilihan nilai *n* tersebut didasarkan pada pengujian manual pada Tabel 5.5 pada Sub Bab Analisis Rencana Pengujian. Persamaan yang digunakan untuk mencari nilai *n-gram* terdapat pada Persamaan 2.1. Dalam persamaan tersebut membutuhkan kode ASCII dari tiap karakter yang akan dilakukan perhitungan. Berikut merupakan kode ASCII dari huruf a sampai z yang dijabarkan dalam bentuk Tabel 3.1.

Tabel 3. 1 Tabel ASCII

No	Karakter	Nilai ASCII	Keterangan
1	a	97	Huruf a latin kecil
2	b	98	Huruf b latin kecil
3	c	99	Huruf c latin kecil
4	d	100	Huruf d latin kecil
5	e	101	Huruf e latin kecil
6	f	102	Huruf f latin kecil
7	g	103	Huruf g latin kecil
8	h	104	Huruf h latin kecil
9	i	105	Huruf i latin kecil
10	j	106	Huruf j latin kecil
11	k	107	Huruf k latin kecil
12	l	108	Huruf l latin kecil
13	m	109	Huruf m latin kecil
14	n	110	Huruf n latin kecil
15	o	111	Huruf o latin kecil
16	p	112	Huruf p latin kecil
17	q	113	Huruf q latin kecil
18	r	114	Huruf r latin kecil
19	s	115	Huruf s latin kecil

No	Karakter	Nilai ASCII	Keterangan
20	t	116	Huruf t latin kecil
21	u	117	Huruf u latin kecil
22	v	118	Huruf v latin kecil
23	w	119	Huruf w latin kecil
24	x	120	Huruf x latin kecil
25	y	121	Huruf y latin kecil
26	z	122	Huruf z latin kecil

Sumber: <https://id.wikipedia.org/wiki/ASCII>

3.3.6 Membentuk Nilai *Hash/Rolling Hash*

Setelah *n-gram* terbentuk dilakukan perhitungan nilai *hash* dari setiap gram. Fungsi yang digunakan untuk menghasilkan nilai *hash* dari rangkaian gram dalam metode *Winnowing* adalah *rolling hash*. Fungsi tersebut merupakan fungsi *hash*, sedangkan nilai yang dihasilkan adalah nilai *hash*. Nilai basis yang digunakan adalah 11 karena besar nilai basis akan mempengaruhi keakuratan persentase yang dihasilkan. Seperti pada penelitian yang dilakukan oleh Maskur dkk. yang menunjukkan bahwa jika menggunakan nilai basis 4 maka rata-rata selisih yang dihasilkan lebih kecil dibandingkan menggunakan nilai basis 2 (Maskur et al., 2020).

3.3.7 Membentuk *Window*

Pada tahap ini akan dilakukan pembentukan *window* dari nilai *hash* di mana nilai *hash* yang telah terbentuk akan dibagi dalam beberapa *window* dengan ukuran *w*. Nilai *window* yang digunakan adalah 7 dan penentuan nilai *window* yang digunakan didasarkan dari hasil perhitungan terbaik antara perhitungan *Winnowing* dengan hasil dari aplikasi Plagiarsm Checker X seperti pada Tabel 5.5 pada Sub Bab Analisis Rencana Pengujian.

3.3.8 Membentuk *Fingerprint*

Dari proses pembentukan *window*, selanjutnya dilakukan pemilihan nilai *hash* terkecil pada tiap *window* untuk dijadikan *fingerprint* tiap dokumen. Alasan diambil nilai *hash* terkecil pada proses *fingerprint* agar dapat mendeteksi penjiplakan termasuk bagian-bagian kecil yang mirip dalam dokumen yang

berjumlah banyak (Ulfa & Mustikasari, 2016). Jika terdapat nilai *hash* terkecil yang sama pada *window* maka yang diambil untuk dijadikan *fingerprint* adalah *window* sebelah kanan.

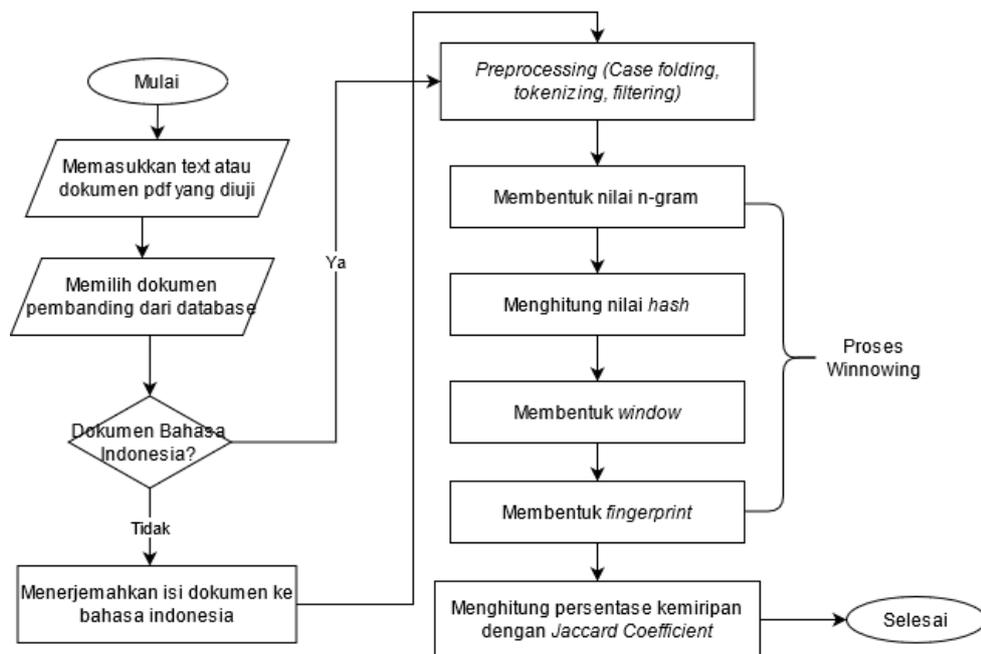
3.3.9 Menghitung Kemiripan Menggunakan *Jaccard Coefficient*

Pada tahap ini berguna untuk menghitung persentase kemiripan antara dokumen uji dengan dokumen pembanding. Nilai *fingerprint* yang dibentuk dari metode *Winnowing* digunakan untuk mengukur persentase kemiripan teks. Persamaan *Jaccard Coefficient* digunakan untuk menghitung nilai kemiripan dari kumpulan kata yang telah dihitung nilai *hash*nya. Untuk menghitung kemiripan dengan *Jaccard Coefficient* dapat menggunakan Persamaan 2.2 yang dituliskan pada Bab 2.

3.4 Desain Sistem

Desain sistem berguna untuk memberikan gambaran tentang rancangan sistem yang akan dibuat. Sistem pendeteksi plagiarisme dokumen teks lintas bahasa secara umum dirancang untuk mendeteksi persentase kemiripan antar dua dokumen dengan bahasa yang berbeda.

3.4.1 Rancangan *Flowchart* Sistem



Gambar 3. 1 *Flowchart* Proses Deteksi Plagiarisme Metode *Winnowing*

Pada Gambar 3.1 merupakan perancangan proses yang ditampilkan dalam bentuk diagram alir atau *flowchart*. Proses yang dilakukan pertama kali untuk mendeteksi plagiarisme adalah harus memasukkan dokumen yang akan diuji dalam bentuk teks atau *file* PDF dan akan disimpan pada *database*. Selain itu pengguna juga dapat membandingkan dokumen uji dengan lebih dari satu dokumen pembanding yang sudah tersimpan di *database*. Teks yang diinputkan akan dideteksi bahasa yang digunakan. Jika teks berbahasa Inggris maka akan diterjemahkan terlebih dahulu ke bahasa Indonesia. Setelah semua teks berbahasa Indonesia, dilakukan proses *preprocessing* untuk menghilangkan karakter atau tanda baca yang tidak perlu pada kalimat dan hanya diambil kata yang penting saja. Setelah itu dilakukan pembentukan nilai *n-gram* dan dilanjutkan dengan menghitung nilai *hash* tiap *n-gram*. Lalu dibentuklah *window* dari nilai *hash* dan setelah itu dilanjutkan dengan pemilihan *fingerprint* dari tiap *window*. Setelah nilai *fingerprint* terbentuk, dilakukan perhitungan nilai kemiripan teks menggunakan persamaan *Jaccard Coeficient*. Hasil proses perhitungan yang dijalankan tersebut akan disimpan dalam *database*. Sehingga dalam *database* akan menyimpan dokumen teks yang diinputkan, nilai *n-gram* dan *w-gram*, hasil terjemahan dari dokumen teks, hasil *preprocessing* dan persentase kemiripan dokumen.

3.4.2 Analisis Kebutuhan Sistem dan Perangkat

Spesifikasi perangkat lunak dapat dilihat pada Tabel 3.2 sedangkan spesifikasi perangkat keras dapat dilihat pada Tabel 3.3.

Tabel 3. 2 Spesifikasi Perangkat Lunak

No.	Nama Perangkat Lunak
1	Sistem Operasi Windows 7/8/10
2	Visual Studio Code
3	Xampp Control Panel v3.2.2
4	MySql
5	Aplikasi Peramban seperti Firefox, Google Chrome dan internet explorer

Tabel 3. 3 Spesifikasi Perangkat Keras

No	Nama Perangkat Keras
1	Processor AMD Ryzen 5
2	Memory RAM 4GB
3	Kapasitas Hardisk 500GB

3.5 Uji Coba Sistem

Pengujian sistem dilakukan apabila semua tahap perancangan dan implementasi program sudah selesai, hal ini bertujuan untuk mengetahui seluruh sistem yang dibangun berjalan sesuai fungsinya. Dalam pengujian sistem terhadap 3 tahap yaitu pengujian *Black Box*, pengujian tingkat akurasi persentase nilai menggunakan metode *Winnowing* dengan aplikasi Plagiarsm Checker X dan pengujian dengan menggunakan kuesioner.

Pada tahap pengujian *Black Box* dilakukan dengan menguji perangkat lunak tanpa mengetahui struktur internal kode atau program. Sedangkan pada pengujian tingkat akurasi persentase nilai dilakukan dengan menguji dua dokumen teks untuk dicari tingkat kemiripannya menggunakan *Jaccard Coefficient*. Hasil persentase yang didapatkan dari perhitungan *Jaccard Coefficient* akan dibandingkan dengan hasil yang didapatkan dari aplikasi Plagiarism Checker X. Pada pengujian menggunakan kuesioner dilakukan uji coba langsung oleh *user* dan setelah itu *user* dapat mengisi pertanyaan pada kuesioner.