

SKRIPSI

SISTEM DETEKSI PLAGIARISME SKRIPSI
MENGUNAKAN METODE ALGORITMA *WINNOWER*



SUHARSO ARDHI SAPUTRO
NPM. 17.0504.0079

PROGRAM STUDI TEKNIK INFORMATIKA S1
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH MAGELANG
AGUSTUS,2021

BAB 1

PENDAHULUAN

1.1 Latar Belakang Permasalahan

Pesatnya perkembangan teknologi informasi seperti internet menyebabkan banyaknya informasi yang tersedia. Hal ini memudahkan seseorang dalam melakukan penjiplakan suatu karya khususnya karya tulis seperti skripsi. Skripsi suatu proses yang harus dilaksanakan dan diselesaikan dalam suatu jenjang pendidikan untuk memperoleh gelar sebagai seorang sarjana. Dalam menyelesaikan suatu skripsi maka perlu adanya proses yang harus dilaksanakan agar semuanya dapat berjalan secara sistematis sesuai panduan yang digunakan. Namun permasalahan dikalangan mahasiswa tingkat akhir yaitu terkait dengan budaya plagiarisme. Permasalahan ini sebelumnya telah dikaji oleh Fajar Arista tentang plagiarisme marak terjadi di kalangan mahasiswa dan dikategorikan sebagai tindakan pelanggaran karena mengambil karya orang lain (Arista & Listyani, 2015).

Dalam dunia akademik di Universitas masih banyak ditemukan plagiarisme yang disebabkan oleh keterbatasan manusia (Arista & Listyani, 2015). Permasalahan ini juga terjadi di Universitas Muhammadiyah Magelang khususnya prodi Teknik informatika S-1. Sistem yang berjalan saat ini menggunakan dokumen berupa borang skripsi. Borang skripsi yang digunakan terdiri dari latar belakang, rumusan masalah, tujuan, manfaat, penelitian relevan dan landasan teori dalam menilai tingkat kemiripan atau plagiarisme pada sebuah borang skripsi. Namun demikian penelitian ini difokuskan untuk latar belakang pada borang skripsi yang diajukan oleh mahasiswa. Penilaian latar belakang pada borang skripsi yang dilakukan saat ini masih menggunakan persepsi dosen. Prodi Teknik informatika S-1 belum mempunyai sistem yang mampu membandingkan antar skripsi dan mengetahui tingkat persentase kemiripan borang skripsi dengan skripsi terdahulu. Sedangkan, skripsi di Prodi Teknik Informatika Universitas Muhammadiyah Magelang pada tahun 2014-

2019 ditemukan sebanyak 356 skripsi, sehingga rentan terjadinya plagiarisme pada skripsi yang diajukan oleh mahasiswa.

Plagiarisme merupakan suatu hal yang tidak terpuji karena mengakui karya orang lain menjadi miliknya sendiri tanpa mencantumkan sumbernya (Sunardi et al., 2018), (Sinaga, 2018). Pada perkembangan teknologi industri 4.0 dan society 5.0 maka segala bentuk kecurangan dapat diketahui dengan berbagai aplikasi pendeteksi kemiripan teks skripsi yang digunakan sebelumnya. Saat ini penerapan suatu algoritma untuk mendeteksi kemiripan sedang banyak dipergunakan suatu Perguruan Tinggi di Indonesia. Hal tersebut dilakukan untuk mencegah plagiarisme yang dilakukan di kalangan mahasiswa. Dalam penelitian ini Sistem yang menggunakan Algoritma Winnowing dan jaccard similarity. Algoritma Winnowing dipergunakan untuk mendeteksi adanya kemiripan kata pada dua skripsi penelitian yang akan dibandingkan dengan konsep: penghapusan karakter, *n gram*, *hashing*, *window* dan *document fingerprints* (Alamsyah et al., 2019). Sedangkan *Jaccard similarity* merupakan metode yang berfungsi untuk menghitung nilai kemiripan dengan membandingkan dua sampel (Nurdiana et al., 2016). Penelitian yang dilakukan oleh Wibowo tentang penerapan algoritma pada dokumen tugas akhir mahasiswa dengan menggunakan metode *n grams winnowing* dan *jaccard coefficient*, di mana diperoleh hasil penelitian bahwa metode tersebut mampu mendeteksi kemiripan teks dengan hasil yang akurat (Wibowo & Hastuti, 2016).

Berdasarkan penjelasan di atas dapat disimpulkan bahwa diperlukan sebuah sistem yang mampu untuk membandingkan sejauh mana nilai kemiripan dari beberapa skripsi yang sudah ada. Sehingga penelitian ini diharapkan dapat meminimalisir dan mengetahui plagiarisme pada skripsi. Sistem ini akan melakukan pengecekan pada latar belakang skripsi, peneliti memilih latar belakang skripsi dikarenakan dalam latar belakang terkandung identifikasi masalah, metode yang digunakan untuk menyelesaikan masalahnya serta terdapat ide dari penulis yang akan diajukan. Sistem yang diajukan bisa mendeteksi file berekstensi .pdf. kemudian skripsi yang diujikan bersumber dari

borang mahasiswa S-1 dan pembandingnya bersumber dari skripsi publikasi. Dalam sistem yang dibuat akan di tambahkan output surat keterangan plagiarisme sebelum mahasiswa melakukan bimbingan, surat tersebut sebagai penguat bahwa skripsi yang akan diajukan telah bebas dari plagiarisme dan sebagai pendukung bahwa ini skripsi hasil sendiri.

1.2 Rumusan Masalah

Berdasarkan latar belakang permasalahan di atas, maka rumusan masalah penelitian ini yaitu, Bagaimana menerapkan algoritma *winnowing* pada sistem pendeteksi plagiarisme pada latar belakang pada borang skripsi?

1.3 Tujuan Penelitian

Tujuan penelitian yang akan dicapai yaitu: Mengetahui persentase plagiarisme pada latar belakang borang skripsi di Prodi Teknik Informatika Universitas Muhammadiyah Magelang.

1.4 Manfaat Penelitian

Manfaat penelitian yang diharapkan apabila tujuan penelitian tercapai adalah:

1. Memudahkan tim skripsi untuk menyeleksi skripsi mahasiswa Prodi Teknik Informatika Universitas Muhammadiyah Magelang yang diajukan.
2. Mengurangi plagiarisme skripsi di Prodi Teknik Informatika Universitas Muhammadiyah Magelang.

BAB 2

TINJAUAN PUSTAKA

2.1 Penelitian Relevan

Penelitian yang berjudul “Implementasi Deteksi *Plagiarisme* Menggunakan Metode *n-gram* dan *Jaccard Similarity* terhadap Algoritma *Winnowing*” terdapat permasalahan plagiarisme terhadap kalimat. Dengan adanya permasalahan di atas, Yudhana dkk mengatasi menggunakan metode *n-gram* dan *jaccard similarity* terhadap algoritma *winnowing*, di mana diperoleh hasil dari penelitian ialah *n-gram* sangat memengaruhi hasil similarity dan penggunaan *n-gram* yang tepat sangat diperlukan. Jika nilai *n-gram* dan *k-gram* yang dimasukkan semakin kecil maka hasil persentase kesamaanya semakin tinggi itupun berlaku sebaliknya. *Jaccard similarity* yang digunakan menghasilkan tingkat kesamaan antar dokumen juga cukup baik karena koefisien ini sederhana dengan mencari item yang sama dari dua dokumen, kemudian dibagi dengan total item kedua dari penggabungan dua dokumen. Pendeteksi plagiarisme menggunakan metode *n-gram* dan *jaccard similarity* terhadap algoritma *winnowing* cukup baik untuk membandingkan kesamaan antara dua dokumen dan cukup baik digunakan untuk meminimalisasi tindakan plagiarisme (Yudhana *et al.*, 2018).

Penelitian yang berjudul “Aplikasi Pendeteksi Tingkat Kesamaan Dokumen Teks: Algoritma *Rabin Karp* Vs *Winnowing*”. Tindakan *copy paste* dokumen teks sering terjadi dalam penulisan karya ilmiah tanpa memberikan manfaat kepada penulis dokumen teks tersebut. Penelitian ini menggunakan metode algoritma *winnowing* dan algoritma *rabin-karp*. Hasil dari penelitian Sugiyono dkk yaitu perbandingan antara algoritma *rabin-karp* dengan algoritma *winnowing* berdasarkan akurasi dan waktu pemrosesan. Algoritma *winnowing* lebih unggul dari tingkat akurasi dan waktu pemrosesan. Algoritma *rabin-karp* memiliki akurasi sebesar 37,5% dengan waktu proses selama 0,19 detik, sedangkan algoritma *winnowing* memiliki tingkat akurasi sebesar 88,89%

dengan waktu proses selama 0,13 detik. Membuktikan bahwa algoritma *rabin-karp* dan algoritma *winnowing* mampu mendeteksi tingkat kesamaan teks dengan level yang berbeda dapat disimpulkan bahwa Algoritma *winnowing* lebih unggul dibandingkan dengan algoritma *rabin-karp* (Sugiono et al., 2018).

Penelitian yang berjudul “*Text Documents Plagiarisme Detection using Rabin-Karp and Jaro-Winkler Distance Algorithms*”. Dalam penelitian tersebut mengungkapkan semakin banyak kemudahan yang diberikan kepada pengguna, tentu ada masalah yang mengganggu salah satunya adalah plagiarisme pada dokumen teks, penelitian menggunakan algoritma *rabin-karp* dengan *jaro-winkler* untuk menyelesaikan plagiarisme pada dokumen teks. Di mana hasil menyatakan bahwa rata-rata kemiripan *rabin-karp* 51% waktu kurang dari 0,594 menit sedangkan *jaro-winkler* kemiripan 35% dan waktu rata-rata lebih dari 0,992 menit. Sehingga dapat disimpulkan bahwa *rabin-karp* lebih efektif daripada dengan algoritma *jaro-winkler* (Leonardo & Hansun, 2017).

Penelitian dengan judul “*Implementasi Algoritma Jaro-Winkler Distance* untuk Sistem Pendeteksi Plagiarisme pada Dokumen Skripsi”. Dalam penelitiannya untuk mengetahui seberapa besar persentase kesamaan bisa menggunakan pendekatan *string metric*, yaitu dengan melakukan perbandingan *string* menggunakan algoritma *jaro-winkler distance* dengan tahapan melakukan proses menghitung panjang *string*, menentukan jumlah karakter yang sama, menghitung nilai transposisi, menghitung nilai *jaro-distance*, menghitung nilai *jaro-winkler distance*, dan menghitung tingkat kemiripan (*similarity*). Sehingga hasil dari penelitian yang dilakukannya ialah sistem pendeteksi plagiarisme dengan menggunakan algoritma *Jaro-Winkler Distance* dapat digunakan untuk mendeteksi plagiarisme dokumen skripsi dengan cara melakukan perbandingan antara dokumen asli dan dokumen uji yang diinputkan untuk mengetahui tingkat kemiripan (*similarity*) dari dokumen skripsi yang diuji (Novantara, 2018) .

Penelitian yang berjudul “*Sistem Pendeteksi Dini Plagiarisme Menggunakan Algoritma Levenshtein Distance*”. Permasalahan yang diungkapkan oleh peneliti yaitu masih melakukan pendeteksi dokumen yang memiliki indikasi plagiarisme secara manual. Oleh karena itu peneliti

mengusulkan sebuah sistem pendeteksi plagiarisme pada dokumen teks yang dilakukan secara kompatibel. Penelitian tersebut menggunakan algoritma levenshtein distance untuk mendeteksi plagiarisme pada dokumen teks. Hasil menggunakan algoritma levenshtein distance, Pada pengujian menggunakan data *real* dokumen yang memiliki perbedaan dari dokumen 1 dengan *count words* berjumlah 629 dan dokumen 2 dengan *counts words* 128 diketahui nilai kesamaan 13,50% dan nilai perbedaan 86,50%. Pada pengujian efektivitas kecepatan perhitungan, dokumen sumber dengan *count words* 4405 diuji dengan 3 dokumen pembanding dengan total *count words* 13465 memiliki kecepatan menghitung 3,57 detik. Kecepatan tergantung pada jumlah total *words* yang terkandung dalam dokumen sumber dan dokumen pembanding. (Febiawan et al., 2020).

Berdasarkan dari penelitian terdahulu tentang sistem pendeteksi plagiarisme karya ilmiah mahasiswa bahwa masih ditemukan beberapa kelemahan dari penggunaan sistem tersebut seperti: proses *input* teks atau dokumen pembanding dan dokumen uji yang digunakan masih menggunakan ekstensi dokumen (.doc), menggunakan satu dokumen pembanding dan waktu yang digunakan untuk analisis *similarity* dokumen masih cukup lama. Penelitian sistem deteksi plagiarisme yang akan dibuat mudah tanpa proses penginputan nilai *n-gram* atau *window*, sistem ini dapat melakukan cek plagiarisme dengan menggunakan algoritma winnowing dengan file pengujian yang ber ekstensi .pdf. Dalam sistem yang dibuat akan di tambahkan output surat keterangan plagiarisme sebelum mahasiswa melakukan bimbingan, surat tersebut sebagai penguat bahwa skripsi yang akan diajukan telah bebas dari plagiarisme dan sebagai pendukung bahwa ini skripsi hasil sendiri.

2.2 Landasan Teori

2.2.1 Sistem

Sistem dapat didefinisikan sebagai kumpulan komponen-komponen yang saling berelasi dan berinteraksi, serta hubungan antar objek bisa dilihat sebagai satu kesatuan yang dirancang untuk mencapai satu tujuan yang telah ditetapkan. Ada beberapa elemen sistem yang dikemukakan oleh (Irianto, 2012) bahwa elemen tersebut yaitu:

a. Tujuan

Setiap sistem memiliki tujuan (*goal*) entah hanya satu atau mungkin banyak, tujuan inilah yang menjadi pemotivasi yang mengarahkan sistem.

b. Masukan

Masukan (*input*) sistem adalah segala sesuatu yang masuk ke dalam sistem dan selanjutnya menjadi menjadi bahan untuk diproses.

c. Proses

Merupakan bagian yang melakukan perubahan atau transformasi dari masukan menjadi keluaran yang berguna.

d. Keluaran

Keluaran (*output*) merupakan hasil pemrosesan. Pada sistem informasi bisa berupa suatu informasi, luaran dapat berupa suatu informasi, saran, cetakan laporan dan sebagainya.

e. Mekanisme pengendalian dan umpanbalik

Mekanisme pengendalian (*control mechanism*) diwujudkan dengan menggunakan umpan balik (*feedback*), yang mencuplikan luaran. Umpan balik ini digunakan untuk mengendalikan baik masukan maupun proses.

f. Lingkungan

Lingkungan adalah segala sesuatu yang berada di luar sistem, Lingkungan bisa berpengaruh terhadap operasi sistem dalam arti bisa merugikan atau menguntungkan sistem itu sendiri.

g. Batas

Batas (*boundary*) sistem adalah pemisah antara sistem dan daerah diluar sistem (lingkungan). Batas sistem menentukan konfigurasi, ruang lingkup atau kemampuan sistem.

2.2.2 Plagiarisme

Plagiarisme merupakan proses pengambilan gagasan dari orang lain tanpa menyertakan citasi penulis asli dan membohongi banyak khalayak bahwa karya tersebut sebagai miliknya sendiri. Hal ini dapat dicontohkan dengan menggunakan poin-poin umum atau mengutip beberapa kata dan mengubahnya dari tulisan asli tanpa menyebutkan sumber tulisan (Fadhli, Mulkan Muttaqin, 2020).

Ada beberapa tipe plagiarisme yang oleh (Hermawan, 2019) dijelaskan sebagai berikut:

1. Plagiarisme Langsung (*Direct Plagiarisme*)

Disebut plagiarisme langsung apabila pelaku mengcopy langsung tulisan sebagian atau keseluruhan dan tidak menunjukkan bagian itu sebagai hasil kutipan atau karya orang lain.

2. Plagiarisme Tidak Jelas (*Incorrect Citation*)

Plagiarisme dalam jenis ini manakala pelaku mengutip sebagian karya tulis, tetapi tidak jelas menyebutkan di mana awal kutipan dan di mana akhir kutipan.

3. Plagiarisme Mozaik (*Mosaic Plagiarisme*)

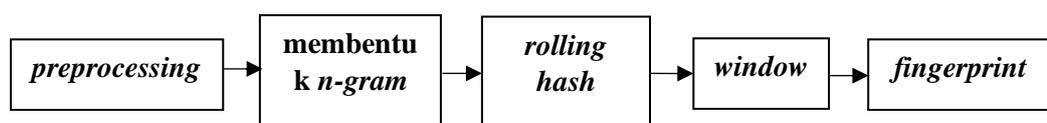
Artinya pelaku mengutip suatu bagian karya tulis dengan mengubah menurut kata-katanya sendiri meskipun yang diubah hanya kata-kata tertentu.

2.2.3 Algoritma *Winnowing*

Salah satu algoritma yang digunakan untuk mendeteksi bentuk kesamaan pada dokumen teks adalah algoritma *winnowing*. Pada dasarnya sistem pendeteksian harus memiliki 3 unsur utama yang harus dipenuhi, seperti (Astutik et al., 2014):

1. *Whitespace insensitivity*, sistem pencocokan teks seharusnya tidak terpengaruh pada spasi, adanya huruf kapital, berbagai tanda baca, dan sebagainya.
2. *Noise surpression*, sistem haruslah menghindari pencocokan kata yang terlalu pendek.
3. *Position independence*, sistem seharusnya tidak bergantung pada posisi kata yang dicari sehingga apabila ditemukan kata yang terindeksi sama dengan posisi berbeda masih dapat dikenal.

Algoritma *winnowing* juga berfungsi sebagai *document fingerprints* atau algoritma yang digunakan untuk mendeteksi tindakan plagiarisme dengan menggunakan teknik *hashing*. *Input* dari algoritma *winnowing* berupa dokumen teks, dan akan menghasilkan keluaran berupa kumpulan nilai *hash* yang terbentuk dari perhitungan ASCII pada setiap karakter. Dan nilai-nilai *hash* yang akan digunakan sebagai *fingerprints* untuk mendeteksi adanya suatu tindakan plagiarisme. Menurut (Wulan et al., 2018) alur dari algoritma *winnowing* itu sendiri yaitu:



Gambar 2. 1 Alur Algoritma *Winnowing*

1. *Preprocessing* adalah proses menghilangkan karakter yang tidak relevan pada dokumen teks, seperti tanda baca, tanda spasi dan mengubah huruf besar menjadi kecil.
2. Membentuk *n-gram* adalah merupakan metode yang digunakan dalam proses tokenisasi atau pemisahan teks, dengan cara membentuk *substring* sepanjang k karakter dari sebuah *string*.
3. *Rolling hash* adalah merupakan teknik yang digunakan untuk mendapatkan nilai *hash* dari rangkaian *grams* yang telah terbentuk dari metode *k-grams* dan memberikan kemampuan untuk menghitung nilai tanpa mengulangi seluruh *string*.
4. *Window* adalah pembagian atau pengelompokan beberapa nilai *hash* dengan ukuran yang ditentukan.
5. *Fingerprints* adalah nilai dari *window* yang telah dibentuk dilakukan pemilihan nilai *hash* terkecil pada tiap *window*.

Nilai *hash* merupakan nilai numerik yang dibentuk dari kode ASCII. Berikut rumus *rolling hash* (Ulfa *et al.*, 2016):

$$H_{(c1...cK)} = c1 * b^{(k-1)} + c2 * b^{(k-2)} + \dots + c^{(k-1)} * b + ck \dots \dots \dots (2.1)$$

Untuk menghitung hash dari k-gram $c2...ck + 1$, dapat menggunakan persamaan (2.2):

$$H_{(c2...ck+1)} = (H_{(c1...ck)} - c1 * b^{(k-1)}) * b + c^{(k+1)} \dots \dots \dots (2.2)$$

Dimana:

$H_{(c1...cK)}$: hashing
c	: nilai ASCII
b	: nilai basis
k	: banyaknya karakter (<i>n-gram</i>)

2.2.4 Jaccard Similarity

Similarity Jaccard kadang disebut *Jaccard Coeficient* yaitu salah satu metode yang membandingkan kesamaan dan keragaman 2 set anggota sampel. Untuk melihat mana yang dibagikan dan mana yang berbeda.

Kesamaan dalam dua bagian atau kumpulan data, dengan rentang dari 0% hingga 100%. Semakin tinggi persentasenya, semakin mirip dua populasi. Meskipun mudah ditafsirkan, tetapi sangat sensitif terhadap ukuran sampel yang kecil dan dapat memberikan hasil yang salah, terutama dengan sampel yang sangat kecil atau kumpulan data dengan pengamatan yang hilang (Fitrianto & Wibisono, 2018).

$$D A, B = \frac{|A \cap B|}{|A \cup B|} \times 100 \dots\dots\dots (2.3)$$

Keterangan:

A: nilai *fingerprints* dokumen 1

B: nilai *fingerprints* dokumen 2

2.2.5 Basis Data

Basis data (*database*) merupakan perpaduan data yang saling berkaitan dalam satu tempat atau bisa dibidang ‘wadah’ agar data bisa diolah, untuk memperoleh kemudahan, ketepatan dan kecepatan ketika pengambilan data kembali. (Suhartati & Dwi Atma, 2017).

Terdapat beberapa elemen basis data, yaitu:

- a. *Database* adalah kumpulan tabel yang mempunyai kaitan antara suatu tabel dengan tabel lainnya sehingga membentuk suatu bangunan data.
- b. Tabel adalah kumpulan *record* yang mempunyai panjang elemen yang sama dan atribut yang sama namun berbeda data *valuenya*.
- c. Entitas adalah sekumpulan objek yang terdefiniskan yang mempunyai karakteristik sama dan bisa dibedakan satu dengan lainnya. Objek dapat berupa barang, orang, tempat atau suatu kejadian.
- d. Atribut adalah deskripsi data yang bisa mengidentifikasi entitas yang membedakan entitas dengan entitas yang lain. Seluruh

atribut harus cukup untuk menyatakan identitas objek atau dengan kata lain, kumpulan atribut dari setiap entitas dapat mengidentifikasi keunikan suatu individu.

- e. Data *Value* (Nilai Data) adalah data aktual atau informasi yang disimpan pada tiap data, elemen atau atribut. Atribut nama pegawai menunjukkan tempat di mana informasi nama karyawan disimpan, nilai datanya misalnya adalah Anjang, Arif, Suryo dan lain-lain yang merupakan isi data nama pegawai tersebut.
- f. File adalah kumpulan *record* sejenis yang mempunyai panjang elemen yang sama, atribut yang sama namun berbeda nilai datanya.
- g. *Record/Tuple* adalah kumpulan elemen-elemen yang saling berkaitan menginformasikan tentang suatu entitas secara lengkap. Satu *record* mewakili satu data atau informasi (Irsyad, 2017).

2.2.6 PHP (*Hypertext Preprocessor*)

PHP mempunyai kepanjangan yaitu *Hypertext Preprocessor* merupakan suatu bahasa pemrograman yang difungsikan untuk membangun suatu site dinamis. PHP menyatu dengan kode HTML, maksudnya adalah beda kondisi. HTML digunakan sebagai pembangun atau fondasi dari kerangka format web, sedangkan PHP difungsikan sebagai prosesnya. Adanya PHP tersebut, sebuah web akan sangat mudah *dimaintenance*. PHP berjalan pada sisi server sehingga PHP disebut juga sebagai bahasa *Server Side Scripting*. Artinya bahwa dalam setiap/untuk menjalankan PHP wajib adanya web server (Kamran & Kharie, 2016).

2.2.7 UML (*Unified Modelling Language*)

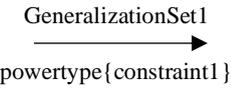
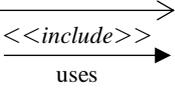
Pada perkembangan teknik pemrograman berorientasi objek, muncullah sebuah standarisasi bahasa pemodelan untuk pembangunan

perangkat lunak yang dibangun dengan menggunakan teknik pemrograman berorientasi objek yaitu *Unified Modeling Language* (UML). UML adalah salah satu standar bahasa yang banyak digunakan di dunia industri untuk mendefinisikan kebutuhan, membuat analisis & desain, serta menggambarkan arsitektur dalam pemrograman berorientasi objek (Putra & Andriani, 2019). UML memiliki elemen grafis yang bisa dikombinasikan menjadi sebuah diagram. Hasil diagram tersebut akan menggambarkan atau mendokumentasikan beberapa aspek dari sebuah sistem (Oktaviyani et al., 2018). Diagram yang ada dalam UML antara lain:

a. *Usecase Diagram*

Merupakan pemodelan untuk kelakuan sistem informasi yang akan dibuat (Fauzi & Retnoningsih, 2018). Diagram ini menjelaskan sebuah interaksi antara satu atau lebih aktor dengan sistem informasi yang akan dibuat.

Tabel 2. 1 *Usecase Diagram* (Aprianti & Maliha, 2016)

Simbol	Deskripsi
	Fungsionalitas yang disediakan sistem sebagai unit-unit yang saling bertukar pesan antar unit dan aktor
 <i>Actor</i>	Orang, proses, atau sistem lain yang berinteraksi dengan sistem informasi yang akan dibuat diluar sistem informasi
 Asosiasi	Relasi antar kelas dengan makna umum, asosiasi biasanya disertai dengan <i>multiplicity</i>
	Hubungan generalisasi dan spesialisasi antara dua buah <i>Use Case</i>
	Relasi <i>Use Case</i> tambahan ke sebuah <i>Use Case</i> dimana <i>Use Case</i> yang ditambahkan memerlukan <i>Use Case</i> untuk menjalankan fungsinya
	Relasi <i>Use Case</i> tambahan ke sebuah <i>Use Case</i> dimana <i>Use Case</i> yang ditambahkan memerlukan <i>Use Case</i> untuk menjalankan fungsinya

b. *Activity Diagram*

Merupakan gambaran aliran kerja atau aktivitas dari sebuah sistem atau menu yang ada pada sistem (Destiningrum & Adrian, 2017). Diagram ini menggambarkan aktivitas sistem bukan aktivitas yang dilakukan oleh aktor.

Tabel 2. 2 *Activity Diagram* (Aprianti & Maliha, 2016)

Simbol	Deskripsi
	Status awal aktivitas pada sebuah diagram aktivitas
	Aktivitas yang dilakukan sistem, aktivitas biasanya diawali dengan kata kerja
	Asosiasi percabangan di mana jika ada pilihan aktivitas lebih dari satu
	Asosiasi penggabungan di mana lebih dari satu aktivitas digabungkan menjadi satu

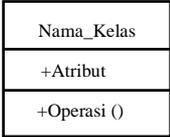
c. *Sequence Diagram*

Sequence Diagram menggambarkan kelakuan objek pada *UseCase* dengan mendeskripsikan waktu hidup objek dan *message* yang dikirimkan dan diterima antar objek.

d. *Class Diagram*

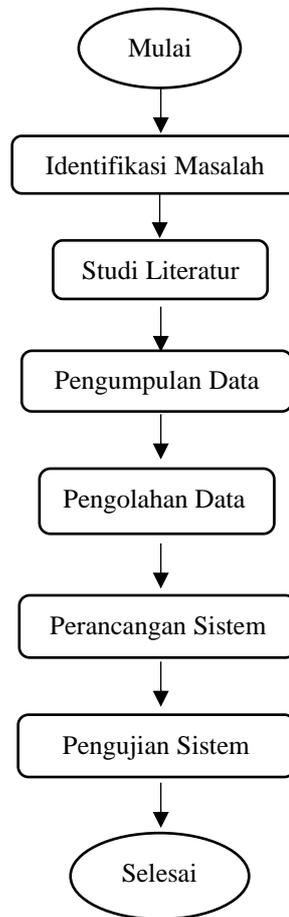
Menggambarkan struktur sistem dari segi pendefinisian kelas-kelas yang akan dibuat untuk membangun sistem. Kelas memiliki apa yang disebut atribut dan metode atau operasi.

Tabel 2. 3 Tabel Notasi *Class Diagram* (Aprianti & Maliha, 2016)

Simbol	Deskripsi
 <p>Kelas</p>	Kelas pada struktur sistem
 <p>Antar Muka</p>	Sama dengan konsep <i>interface</i> dalam pemrograman berorientasi objek
 <p>Asosiasi</p>	Relasi antar kelas dengan makna umum, asosiasi biasanya disertai dengan <i>multiplicity</i>
 <p>Asosiasi Berarah</p>	Relasi antar kelas dengan makna kelas yang digunakan oleh kelas lain, asosiasi biasanya juga disertai dengan <i>multiplicity</i>
 <p>Generalisasi</p>	Relasi antar kelas dengan makna generalisasi-spesialisasi (umum khusus)
 <p>Kebergantungan</p>	Relasi antar kelas dengan makna kebergantungan antar kelas
 <p>Agregasi</p>	Relasi antar kelas makna semua-bagian

BAB 3 METODOLOGI PENELITIAN

Prosedur penelitian dalam penelitian terbagi menjadi 6 langkah, yaitu sebagai berikut:



Gambar 3. 1 Prosedur penelitian

Identifikasi masalah pada penelitian ini yaitu dalam tahap ini peneliti menemukan permasalahan yang akan diteliti. Permasalahan yang diteliti yaitu masih belum adanya sistem pendeteksi skripsi mahasiswa di Prodi Teknik Informatika Universitas Muhammadiyah Magelang (UNIMMA).

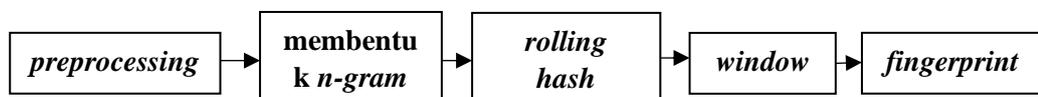
Studi literatur dalam tahap ini peneliti mencari referensi penelitian terdahulu untuk penulisan skripsi dan mencari sumber yang sesuai dan valid, seperti

buku, jurnal dan prosiding serta skripsi. Kemudian peneliti akan mencari metode yang sesuai untuk menyelesaikan masalah yang peneliti temukan.

Pada tahap pengumpulan data peneliti melakukan analisa kebutuhan data yang diperlukan. Kemudian peneliti mengumpulkan data yang dibutuhkan tersebut untuk proses penelitian seperti data mahasiswa, data tim skripsi dan skripsi mahasiswa. Metode pengumpulan data berupa suatu permintaan tentang sifat, keadaan, kegiatan tertentu dan sejenisnya. Pengumpulan data dilakukan untuk mendapatkan suatu informasi yang dibutuhkan dalam mencapai tujuan penelitian. Penelitian ini menggunakan metode:

1. Wawancara, merupakan salah satu teknik pengumpulan data yang dilakukan melalui tatap muka langsung dengan narasumber dengan cara tanya jawab langsung. Wawancara dilakukan dengan Dosen dan Pegawai TU Fakultas Teknik serta Pegawai Perpustakaan Universitas Muhammadiyah Magelang yang berhubungan dengan data yang terkait.
2. Observasi, suatu metode pengumpulan data yang dilakukan dengan mengamati langsung, melihat dan mengambil suatu data yang dibutuhkan di tempat penelitian dilakukan. Observasi juga bisa diartikan sebagai proses yang kompleks. Pengumpulan data yang dilakukan di Prodi Teknik Informatika Universitas Muhammadiyah Magelang (UNIMMA).

Tahap pengolahan data dalam penelitian ini menggunakan algoritma *winnowing*, dan untuk menghitung kemiripannya dengan *jacard similarity*. Dimana nanti akan muncul hasil persentase kemiripan antara data pembanding dengan data sampel. Teknik pengolahan data dalam penelitian ini menggunakan algoritma *winnowing*, dimana algoritma *winnowing* memiliki alur sebagai berikut:



Gambar 3. 2 Alur Metode *winnowing*

1. *Preprocessing* adalah proses menghilangkan karakter yang tidak relevan pada dokumen teks, seperti tanda baca, tanda spasi dan mengubah huruf besar menjadi kecil.
2. Membentuk *n-gram* adalah merupakan metode yang digunakan dalam proses tokenisasi atau pemisahan teks, dengan cara membentuk *substring* sepanjang k karakter dari sebuah *string*.
3. *Rolling hash* merupakan teknik yang digunakan untuk mendapatkan nilai *hash* dari rangkaian *grams* yang telah terbentuk dari metode *k-grams*.
4. *Window* adalah pembagian atau pengelompokan beberapa nilai *hash* dengan ukuran yang ditentukan.
5. *Fingerprints* adalah nilai dari *window* yang telah dibentuk dilakukan pemilihan nilai *hash* terkecil pada tiap *window*.

Algoritma *Winnowing* termasuk algoritma yang memiliki 3 unsur utama dalam pada pendeteksi yaitu:

1. *Whitespace insensitivity*, sistem pencocokan teks seharusnya tidak terpengaruh pada spasi, adanya huruf kapital, berbagai tanda baca, dan sebagainya.
2. *Noise surpression*, sistem seharusnya menghindari pencocokan kata yang terlalu pendek.
3. *Position independence*, sistem seharusnya tidak bergantung pada posisi kata yang dicari sehingga apabila ditemukan kata yang terindeksi sama dengan posisi berbeda masih dapat dikenal.

Pada tahap perancangan sistem, peneliti merancang dan membangun sistem sesuai kebutuhan dan sesuai dengan yang diusulkan dalam penelitian serta memberi kemudahan bagi pengguna nantinya.

Pengujian dalam penelitian ini dilakukan dengan tiga cara yaitu pengujian pada pengguna, kestabilan hasil, dan pengujian perbandingan perhitungan. Pengujian yang dilakukan pada pengguna, dilakukan dengan pengisian kuesioner oleh pengguna berdasarkan penilaian hasil demo sistem yang dilakukan secara langsung. Kemudian untuk menguji tingkat kestabilan hasil plagiarisme, dilakukan dengan beberapakali uji coba pada judul yang sama. Setelah itu peneliti menguji perbandingan perhitungan dengan cara membandingkan hasil perhitungan manual menggunakan excel dengan hasil perhitungan sistem. Peneliti akan menggunakan naskah publikasi sejumlah 30 naskah yang terfokus pada latar belakang dan 2 file latar belakang pada borang skripsi yang akan dicek persentase plagiarisme.

BAB 6

KESIMPULAN DAN SARAN

6.1 Kesimpulan

1. Algoritma *winnowing* dapat diterapkan untuk mendeteksi plagiarisme latar belakang borang skripsi di prodi Teknik Informatika Universitas Muhammadiyah Magelang
2. Sistem yang dibuat dengan penerapan algoritma *winnowing* dapat mengetahui presentase plagiarisme dengan tingkat akurasi 94,93%.
3. Sistem dapat memberikan hasil luaran berupa surat keterangan plagiarisme sebagai media evaluasi skripsi untuk mahasiswa prodi Teknik Informatika Universitas Muhammadiyah Magelang.

6.2 Saran

Perlu adanya pengembangan lebih lanjut, mahasiswa dapat login ke sistem sehingga bisa menggunakan pesan pop up yang dikirimkan kepada mahasiswa untuk mengetahui hasil persentase plagirisme borang skripsi, sehingga tidak menggunakan surat keterangan plagiarisme untuk menginformasikan hasil persentase plagiarisme. Selain itu perlu dikaji ulang untuk mengembangkan sistem saat proses unggah borang skripsi.

DAFTAR PUSTAKA

- Alamsyah, N., Rasyidan, M., Informasi, F. T., Islam, U., Muhammad, K., Al, A., Banjarmasin, B., Informasi, F. T., Islam, U., Muhammad, K., Al, A., & Banjarmasin, B. (2019). *Deteksi plagiarisme tingkat kemiripan judul skripsi pada fakultas teknologi informasi menggunakan algoritma winnowing*. 10(4), 197–201.
- Aprianti, W., & Maliha, U. (2016). *Sistem Informasi Kepadatan Penduduk Kelurahan Atau Desa Studi Kasus Pada Kecamatan Bati-Bati*. 2(2013), 21–28.
- Arista, R. F., & Listyani, R. H. (2015). Rasionalitas Plagiarisme Di Kalangan Mahasiswa. *E Journal UNESA*, 3, 1–5.
- Astutik, S., Cahyani, A. D., & Sophan, M. K. (2014). Sistem Penilaian Esai Otomatis Pada E-Learning Dengan Algoritma Winnowing. *Jurnal Informatika*, 12(2), 47–52. <https://doi.org/10.9744/informatika.12.2.47-52>
- Destiningrum, M., & Adrian, Q. J. (2017). Sistem Informasi Penjadwalan Dokter Berbasis Web Dengan Menggunakan Framework Codeigniter (Studi Kasus: Rumah Sakit Yukum Medical Centre). *Jurnal Teknoinfo*, 11(2), 30. <https://doi.org/10.33365/jti.v11i2.24>
- Fadhli, Mulkan Muttaqin, J. S. (2020). *Panduan belajar manajemen referensi dengan Mendeley* (T. L. Bong (ed.)). Yayasan Kita Menulis.
- Febiawan, M. H., Setiawan, A., & Primadewi, A. (2020). Sistem Pendeteksi Dini Plagiarisme Menggunakan Algoritma Levenshtein Distance. *Jurnal Komtika (Komputasi Dan Informatika)*, 3(1), 18–27. <https://doi.org/10.31603/komtika.v3i1.3464>
- Fitrianto, N., & Wibisono, S. (2018). Sistem Pakar Penanganan Gangguan Layanan Indihome Pada Pelanggan Pt Telkom Indonesia Menggunakan Metode Case-Based. *Prosiding SINTAK 2018*, 472–479.
- Hermawan, A. (2019). Kebijakan Dosen Mengurangi Plagiarisme pada Karya Ilmiah Mahasiswa. *IJIP : Indonesian Journal of Islamic Psychology*, 1(2), 264–284. <https://doi.org/10.18326/ijip.v1i2.264-284>
- Kamran, F., & Kharie, N. F. (2016). Sistem Informasi Akademik Berbasis Web Pada Politeknik Sains Dan Teknologi Wiratama Maluku Utara. *IJIS - Indonesian Journal On Information System*, 1(2), 89–98. <https://doi.org/10.36549/ijis.v1i2.14>
- Leonardo, B., & Hansun, S. (2017). Text documents plagiarism detection using Rabin-Karp and Jaro-Winkler distance algorithms. *Indonesian Journal of*

- Electrical Engineering and Computer Science*, 5(2), 462–471.
<https://doi.org/10.11591/ijeecs.v5.i2.pp462-471>
- Novantara, P. (2018). Implementasi Algoritma Jaro-Winkler Distance Untuk Sistem Pendeteksi Plagiarisme Pada Dokumen Skripsi. *Buffer Informatika*, 3(1). <https://doi.org/10.25134/buffer.v3i2.960>
- Nurdiana, O., Jumadi, J., & Nursantika, D. (2016). Perbandingan Metode Cosine Similarity Dengan Metode Jaccard Similarity Pada Aplikasi Pencarian Terjemah Al-Qur'an Dalam Bahasa Indonesia. *Jurnal Online Informatika*, 1(1), 59. <https://doi.org/10.15575/join.v1i1.12>
- Oktaviyani, N., Bachtiar, L., Darwan, U., & Sampit, A. (2018). *SISTEM INFORMASI ADMINISTRASI BIMBINGAN BELAJAR PADA ENGLISH SOLUTION SAMPIT BERBASIS WEB*. 03.
- Putra, D. W. T., & Andriani, R. (2019). Unified Modelling Language (UML) dalam Perancangan Sistem Informasi Permohonan Pembayaran Restitusi SPPD. *Jurnal TeknoIf*, 7(1), 32. <https://doi.org/10.21063/jtif.2019.v7.1.32-39>
- Riduwan. (2008). *Skala pengukuran variabel-variabel penelitian*. Alfabeta.
- Sinaga, S. Y. (2018). Penggunaan Aplikasi Turnitin Sebagai Sarana Cek Plagiarisme dalam Layanan Perpustakaan Universitas Ukrida. *BIBLIOTIKA : Jurnal Kajian Perpustakaan Dan Informasi*, 2(2), 123–131. <https://doi.org/10.17977/um008v2i22018p123>
- Sugiono, S., Herwin, H., Hamdani, H., & Erlin, E. (2018). Aplikasi Pendeteksi Tingkat Kesamaan Dokumen Teks: Algoritma Rabin Karp Vs. Winnowing. *Digital Zone: Jurnal Teknologi Informasi Dan Komunikasi*, 9(1), 82–93. <https://doi.org/10.31849/digitalzone.v9i1.1242>
- Suhartati, S., & Dwi Atma, Y. (2017). Optimasi Query Sederhana Guna Kecepatan Query Pada Database Server. *Metik Journal*, 1(1), 13–17.
- Sunardi, Yudhana, A., & Mukaromah, I. A. (2018). Implementasi Deteksi Plagiarisme Menggunakan Metode N-Gram Dan Jaccard Similarity Terhadap Algoritma Winnowing. *Transmisi*, 20(3), 105. <https://doi.org/10.14710/transmisi.20.3.105-110>
- Ulfa, N. F., Mustikasari, M., & Bastian, I. (2016). Pendeteksian tingkat similaritas dokumen berbasis web menggunakan algoritma winnowing. *Konferensi Nasional Teknologi Informasi Dan Komunikasi (KNASTIK)*, November, 194–203.
- Wibowo, R. K., & Hastuti, K. (2016). Penerapan Algoritma Winnowing Untuk Mendeteksi Kemiripan Teks pada Tugas Akhir Mahasiswa. *Techno.Com*, 15(4), 303–311.
- Wulan, F. R. N., Kunaefi, A., & Permadi, A. (2018). Deteksi Plagiasi Dokumen Skripsi Mahasiswa Menggunakan Metode N-Grams Dan Winnowing.

Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer, 9(2), 1021–1032.
<https://doi.org/10.24176/simet.v9i2.2535>

Yudhana, A., Sunardi, & Mukaromah, I. A. (2018). Implementation of winnowing algorithm with dictionary English-Indonesia technique to detect plagiarism. *International Journal of Advanced Computer Science and Applications*, 9(5), 183–189. <https://doi.org/10.14569/IJACSA.2018.090523>