Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB II

LANDASAN TEORI

2.1 *Plagiarism*

Menurut Peraturan Menteri Pendidikan Nasional Republik Indonesia Nomor 17 Tahun 2010 dikatakan: "Plagiat adalah perbuatan sengaja atau tidak sengaja dalam memperoleh atau mencoba memperoleh kredit atau nilai untuk suatu karya ilmiah, dengan mengutip sebagian atau seluruh karya dan atau karya ilmiah pihak lain yang diakui sebagai karya ilmiahnya, tanpa menyatakan sumber secara tepat dengan memadai".

Kamus Besar Bahasa Indonesia (2008) disebutkan: "Plagiat adalah pengambilan karangan (pendapat dan sebagainya) orang lain dan menjadikannya seolah-olah karangan sendiri". Dari defenisi *plagiarism* diatas dapat disimpulkan defenisi *plagiarism* secara umum adalah: "Suatu kegiatan dengan mengutip, menggunakan, mengakui, melakukan parafrase menggunakan karya orang lain tanpa mencantumkan identitas sumber asli dari penelitian tersebut".

Sanksi *Plagiarism* telah dicantumkan dalam undang-undang No.12 tahun 2012 tentang pendidikan tinggi pada pasal 28 ayat 5:

"Gelar akademik, gelar vokasi, atau gelar profesi dinyatakan tidak sah dan dicabut oleh Perguruan Tinggi apabila karya ilmiah yang digunakan untuk memperoleh gelar akademik, gelar vokasi, atau gelar profesi terbukti merupakan hasil jiplakan atau plagiat".

Pasal 42 ayat 3:

"Lulusan Pendidikan Tinggi yang menggunakan karya ilmiah untuk memperoleh ijazah dan gelar, yang terbukti merupakan hasil jiplakan atau plagiat, ijazahnya dinyatakan tidak sah dan gelarnya dicabut oleh Perguruan Tinggi".

Beberapa ruang lingkup yang termasuk kedalam kategori *plagiarism*:

1. Mengutip kata-kata atau kalimat orang lain tanpa menggunakan tanda kutip dan tanpa menyebutkan identitas sumbernya.
2. Menggunakan gagasan, pandangan atau teori orang lain tanpa menyebutkan identitas sumbernya.

3. Menggunakan fakta (Data, Informasi) milik orang lain tanpa menyebutkan identitas sumbernya.
4. Mengakui tulisan orang lain sebagai tulisan sendiri.
5. Melakukan parafrase (mengubah kalimat orang lain ke dalam susunan kalimat sendiri tanpa mengubah idenya) tanpa menyebutkan identitas sumbernya.
6. Menyerahkan suatu karya ilmiah yang dihasilkan dan atau telah dipublikasikan oleh pihak lain seolah-olah sebagai karya sendiri.

Putra (2011) menjelaskan bahwa *plagiarism* terdiri dari beberapa bentuk yang terdiri dari:

1. Plagiat Langsung

   Plagiat jenis ini merupakan plagiat jenis berat dimana seseorang menyalin secara langsung sumber kata demi kata tanpa menunjukkan bahwa itu merupakan kutipan dan sama sekali tidak menyebutkan sumbernya.
2. Plagiat karena kutipannya tidak jelas atau salah kutip

   Plagiat jenis ini yaitu plagiat dimana tidak ditunjukkan sumber rujukan secara jelas dimana awal dan akhirannya dalam suatu karya. Jadi dalam hal ini si plagiat tidak membedakan bagian yang sebenarnya ia kutip.
3. Plagiat Mosaik

   Plagiat jenis ini merupakan plagiat yang sering terjadi. Plagiat jenis ini yaitu dimana penulis tidak menyebutkan sumber aslinya, penulis hanya mengubah kata-kata dalam kutipannya. Istilah ini seperti parafrase dimana dalam membuat parafrase kalimat atau kutipan yang akan ditulis diubah dengan kata-kata sendiri namun bedanya dalam pembuatan parafrase harus dicantumkan sumber aslinya dengan jelas.

H, Soelistyo (2011) menjelaskan ada beberapa tipe *plagiarism* yang sering terjadi diantaranya:

1. *Plagiarism* kata demi kata (*Word for word Plagiarism*). Penulis menggunakan kata-kata penulis lain (persis) tanpa menyebutkan sumbernya.
2. Plagiat atas sumber (*Plagiarism of source*). Penulis menggunakan gagasan orang lain tanpa memberikan pengakuan yang cukup tanpa menyebutkan sumber yang jelas.

Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
© Hak cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau
UIN SUSKA RIAU

3. *Plagiarism* kepengarangan (*Plagiarism of Authorship*). Penulis mengakui sebagai pengarang karya tulis karya orang lain.
4. *Self Plagiarism* termasuk dalam tipe penulis mempublikasikan satu artikel pada lebih dari satu redaksi publikasi dan mendaur ulang karya tulis atau karya ilmiah. Sehingga pembaca akan memperoleh hal baru, yang benar-benar penulis tuangkan pada karya tulis yang menggunakan karya lama.

Pendapat lainnya mengenai tipe-tipe *Plagiarism* dijelaskan oleh Indriati (2015) diantaranya yaitu:

1. Kloning

   Kloning merupakan *plagiarism* dimana penulis mengutip kata demi kata seperti aslinya tanpa menyebutkan sumber rujukan dan mengakui gagasan tersebut sebagai miliknya.
2. *Ctrl-C*

   Ctrl-C adalah perintah singkat untuk menyalin teks yang telah dipilih. Plagiat jenis ini merupakan bentuk plagiat yang memuat sebagian besar teks asli dari sumber tanpa adanya perubahan atau pemaknaannya kembali.
3. Tulisan *Find Replace*

   Tulisan *find-replace* adalah tulisan yang hanya mengubah kata kunci atau mengganti dengan *sinonim* pada kata kunci yang ada dalam tulisan rujukan sehingga esensi inti dari tulisan tetaplah sama.
4. Tulisan *Remix*

   Remix merupakan pemaknaan kembali dari beberapa sumber yang sebenarnya berbeda, tapi ditulis atau digabungkan sehingga seolah-olah menjadi satu pendapat yang selaras dari satu sumber.
5. Tulisan Daur Ulang (*Recycle*)

   Recycle merupakan tindakan dimana seseorang menggunakan kembali gagasan yang pernah ia gunakan sebelumnya dengan merubah sedikit pada karya tersebut untuk kemudian dipublikasikan lagi tanpa menyebutkan bahwa sebenarnya gagasan ini sudah pernah ia gunakan sebelumnya.

Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
© Hak cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau
UIN SUSKA RIAU

6. Tulisan *Hybrid*

Tulisan *hybrid* adalah tulisan yang menggabungkan beberapa sumber rujukan, namun yang ditulis dalam kutipan hanya satu sumber.

7. Tulisan *Mashup*

Tulisan *mashup* yaitu menggabungkan beberapa sumber tanpa memperhatikan hubungan dan kesesuaian dari sumber-sumber yang digunakan tersebut.

8. Tulisan "*404 Eror*"

Kesalahan penulisan "*404 error*" merupakan kesalahan dalam menuliskan sumber rujukan yang dapat diakses di internet sehingga tidak dapat diakses atau sumber tersebut memang sebenarnya tidak ada atau rekayasa sumber.

9. Tulisan *Agregator*

Tulisan *agregator* dicirikan sebagai penulisan pendapat atau argumentasi dengan mengutip sebuah sumber dengan teknik yang tepat, tetapi dalam kutipan ini penulis hampir tidak menuangkan pemikirannya.

10. Tulisan *re-tweet*

Kesalahan tipe *re-tweet* merupakan kesalahan dimana sebenarnya penulis sudah menggunakan teknik penulisan kutipan, tapi penulisan tersebut mirip dengan teks atau sesuai dengan kata-kata asli dari teks yang dikutip.

Ada beberapa alasan kenapa seseorang melakukan tindakan *plagiarism* diantaranya:

1. Kurang memahami bagaimana dan kapan saat melakukan kutipan.
2. Kurangnya seseorang dalam menganalisis sumber referensi yang ada.
3. Terbatasnya waktu dalam menyelesaikan karya ilmiah sehingga mendorong seseorang untuk menggunakan cara cepat yaitu *copy paste* hasil karya orang lain.
4. Kurangnya perhatian dari guru, dosen dan pembimbing akademik terhadap persoalan *plagiarism*.

Beberapa upaya telah dilakukan institusi perguruan tinggi untuk menghindarikan masyarakat akademisnya, dari tindakan *plagiarism* sengaja maupun tidak disengaja. Berikut ini pencegahan dan berbagai bentuk pengawasan

Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
© Hak cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau
UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
© Hak cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau

yang dilakukan antara lain (Peraturan Menteri Pendidikan Nasional Nomor 17 Tahun 2010 pasal 7):

1. Karya mahasiswa (Skripsi, Tesis, Disertasi) dilampiri dengan surat pernyataan bermaterai, yang menyatakan bahwa karya ilmiah tersebut tidak mengandung unsur *plagiarism*.
2. Pimpinan perguruan tinggi berkewajiban mengunggah semua karya ilmiah yang dihasilkan di lingkungan perguruan tingginya, seperti portal garuda atau portal lain yang ditetapkan oleh direktorat pendidikan tinggi.
3. Sosialisasi terkait dengan UU No.19 tahun 2002 tentang Hak Cipta dan Peraturan Menteri Pendidikan Nasional Nomor 17 tahun 2010 kepada seluruh masyarakat akademis.

2.2 Pemrosesan Kata

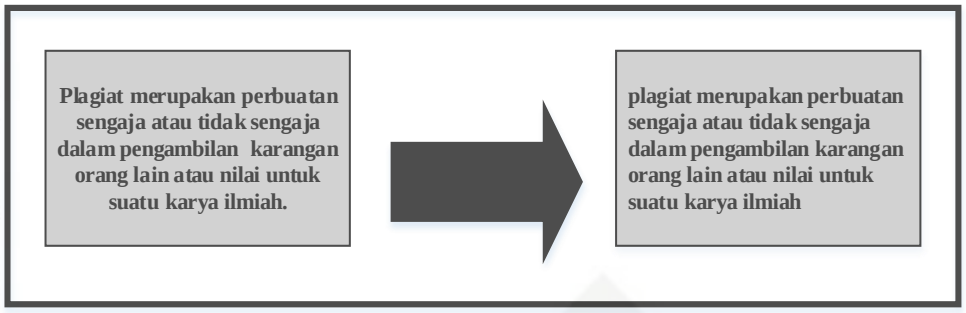
Schutze (2011) Sebelum melakukan perhitungan *similarity* pada suatu dokumen perlu melakukan pra pemprosesan terhadap dokumen kerja praktek agar pendeteksian kemiripan dapat dilakukan dengan cepat dan tepat, sehingga memperoleh hasil yang diinginkan. Tahapan-tahapan pemrosesan kata dapat dilihat sebagai berikut:

a. *Ekstraksi* File

Pada tahapan ini dokumen laporan kerja praktek yang berbentuk PDF, DOC atau DOCX di ubah menjadi file berbentuk TXT. Pada proses ini semua gambar dan tabel yang ada di dalam dokumen ini akan dihilangkan, yang tersisa hanya berupa kata-kata, karena hanya dapat memproses data yang berbentuk teks saja, dengan diubahnya format dokumen kerja praktek menjadi file TXT, maka proses pendeteksian kemiripan kata pada dokumen dapat dilakukan.

b. *Case Folding*

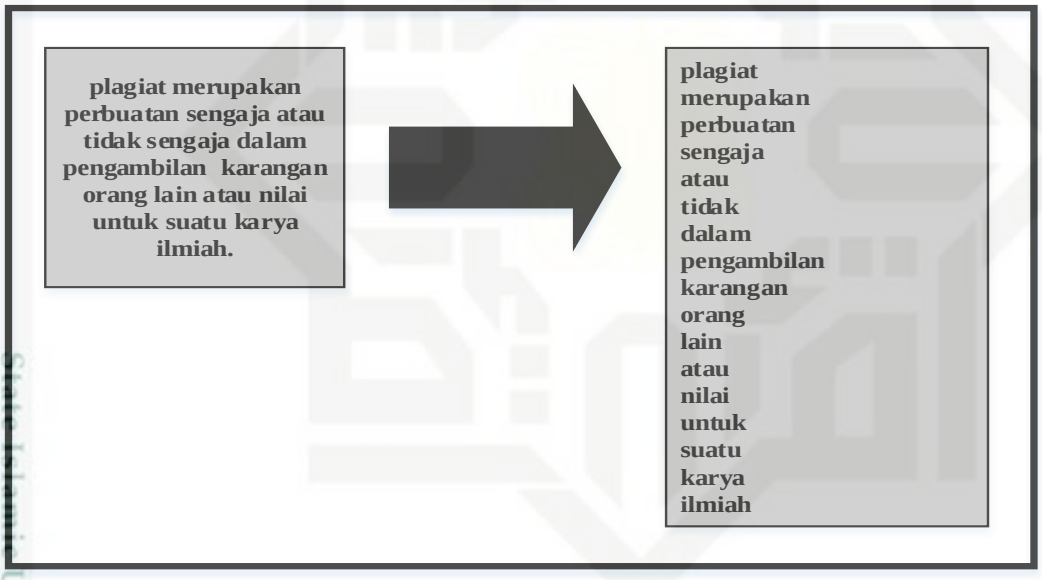
Case folding merupakan proses mengubah semua huruf di dalam dokumen menjadi huruf kecil dan hanya huruf “a” sampai dengan “z” yang diterima. Karakter selain huruf dihilangkan dan dianggap *delimeter*. Contoh dari proses *case folding* dapat dilihat seperti gambar dibawah ini:



Gambar 2.1 *Case Folding*

c. ***Tokenizing***

Tokenizing merupakan pemotongan kalimat menjadi perkata tanpa memperhatikan struktur kalimatnya dan disusun sesuai dengan kalimat aslinya. Contoh dari proses *Tokenizing* dapat dilihat seperti gambar sebagai berikut:



Gambar 2.2 *Tokenizing*

d. *Stopword*

Stopword merupakan proses penghilangan kata-kata umum yang sering muncul dalam sebuah dokumen seperti kata hubung dalam jumlah banyak namun tidak memiliki kaitan di dalam sebuah dokumen, berikut ini contoh-contoh *connectors* (kata hubung):

Tabel 2.1 Kata hubung (*connectors*)

Contoh Kata Hubung			
Atau	Bagaimana	Bilamana	Dengan demikian
Akibatnya	Bagaimanapun juga	Dan	Di mana

Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
© Hak cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau
UIN SUSKA RIAU

Contoh Kata Hubung			
Apakah	Belum lagi	Demikian juga	Jika
Jika tidak	Oleh karena itu	Sebelum	Siapa
Kalau tidak	Oleh karenanya	Sehingga	Tambahan lagi
Ketika	Sama dengan	Sejak, semenjak	Tetapi
Maka dari itu	Sampai	Seperti	Untuk itu
Mengapa	Sebab	Sesudah	Yang
Namun demikian	Sebaliknya	Sewaktu	Yang mana
Berdasarkan	Dalam keadaan	Melalui	Selain dari
Berdasarkan pada	Dalam situasi	Oleh karena	Tanpa memandang
Dalam hal	Dengan, meskipun	Untuk	Sesuai dengan pandangan

Contoh dari *stopword* dapat dilihat sebagai berikut:

Gambar 2.3 *Stopword*

e. *Stemming*

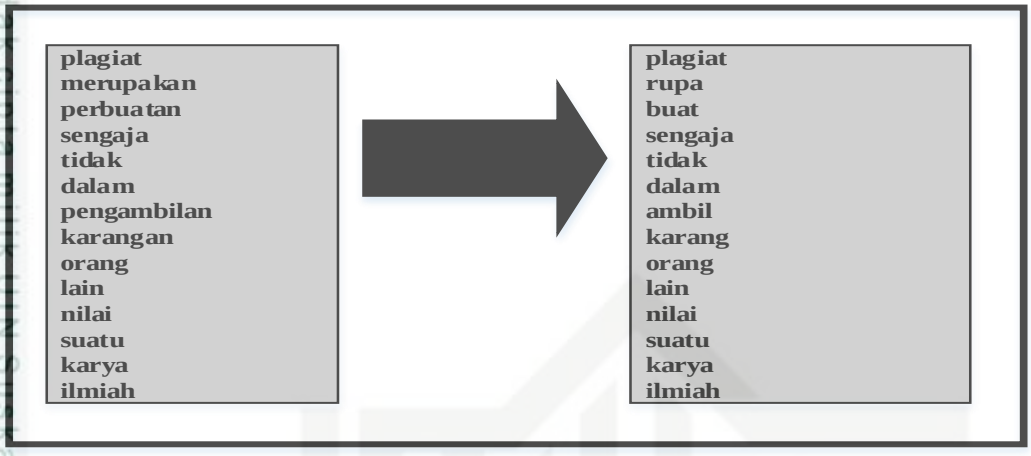
Stemming merupakan proses pengembalian kata dengan menghilangkan imbuhan awal (*prefixs*), imbuhan akhir (*sufixs*), sisipan (*infiks*), dan gabungan imbuhan (*Konfiks*), beberapa imbuhan kata yang sering ditemui di dalam dokumen diantaranya yaitu:

Tabel 2.2 Contoh Imbuhan

Awalan (*perfiks*)	Akhiran (*sufiks*)	Sisipan (*infiks*)	Gabungan imbuhan (*Konfiks*)
Meng-	-an	-el-	Meng-....-kan
Di-	-i	-em-	Meng-....-i
Peng-	-kan	-er-	Peng-....-an
Ber-			Ke-....-an
Ter-			Se-....-nya
Se-			Per-....-an

Contoh dari proses stemming dapat dilihat sebagai berikut:

Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
© Hak Cipta Milik State Islamic University of Sultan Syarif Kasim Riau
UIN SUSKA RIAU



Gambar 2.4 ***Stemming***

2.3 Metode Pembobotan *Term Frequency-Inverse Document Frequency*

Schutze (2011) *Term Frequency* adalah menetapkan bobot (tf_t) pada setiap kata tergantung kemunculan *term* atau kata tersebut dalam suatu dokumen dengan menghitung *term* dalam dokumen "d" berdasarkan berapa banyak kata tersebut mengulang di dokumen "d". *Inverse Document Frequency* adalah jumlah dokumen dalam koleksi yang mengandung *term* tersebut. Jadi *Term Frequency-Inverse Document Frequency* adalah sebuah metode pembobotan *term* di dalam suatu dokumen dengan menghitung kemunculan setiap kata pada dokumen. Pada metode pembobotan ini, perhitungan bobot *term* dalam sebuah dokumen dilakukan dengan mengalikan nilai *Term Frequency* dengan *Inverse Document Frequency.*

a. *Term Frequency*

Term Frequency adalah menetapkan bobot kata pada suatu dokumen berdasarkan jumlah kemunculannya dalam dokumen tersebut, untuk menghitung *Term Frequency* .

b. *Inverse Document Frequency*

Inverse Document Frequency dapat dicari menggunakan persamaan (2.1) dengan rumus dibawah ini:

$$idf_t = \text{Log}\ \frac{N}{df_t} \quad (2.1)$$

Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
© Hak Cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau
UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
© Hak cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau

Keterangan:

a) idf_t adalah bobot *Inverse Document Frequency*

b) N adalah jumlah dokumen yang ada

c) df_t adalah banyaknya dokumen yang mengandung *term* atau kata "t".

c. *Term Frequency-Inverse Document Frequency*

Untuk menghitung bobot tf-idf $_{(t,d)}$ setiap kata "t" di dalam dokumen "d" dicari menggunakan persamaan (2.2) dengan persamaan dibawah ini:

$$\text{tf-idf}_{(t,d)} = \text{tf}_{(t,d)} * \text{idf}_t \quad (2.2)$$

Keterangan:

a) tf-idf $_{(t,d)}$ adalah bobot tf-idf kata "t" di dalam dokumen "d"

b) $tf_{(t,d)}$ adalah jumlah *frequency* kemunculan kata "t" di dalam dokumen "d"

c) idf_t adalah bobot *inverse document frequency*

d. Perhitungan bobot dokumen sumber dan dokumen uji

Perhitungan bobot dokumen sumber dan dokumen uji dapat menggunakan persamaan (2.3) dengan rumus dibawah ini:

$$\vec{V}(d1).\vec{V}(d2) = \sum_{i=1}^{t} (\text{tf} - \text{idf}_{(t,d1)}.\text{tf} - \text{idf}_{(t,d2)}) \quad (2.3)$$

Keterangan:

$\vec{V}(d1)$ = Bobot dokumen d1

$\vec{V}(d2)$ = Bobot dokumen d2

d1 = Bobot *term* i pada dokumen 1

d2 = Bobot *term* i pada dokumen 2

t = Teks

e. Panjang Vektor

Panjang vektor dapat dicari menggunakan persamaan (2.4) dengan rumus dibawah ini:

$$|\vec{V}(d)| = \sqrt{\sum_{i=1}^{t} \text{tf} - \text{idf}_{(t,d)}{}^2} \quad (2.4)$$

Keterangan:

$|\vec{V}(d)|$ adalah panjang vektor bobot dokumen

d adalah bobot kata pada dokumen

f. Persentasi *Similarity*

Similarity merupakan sebuah pegukuran seberapa besar tingkat kemiripan atau kesamaan teks suatu dokumen sumber dengan dokumen uji. Persentasi *similarity* dapat dihitung menggunakan persamaan (2.5) dengan rumus *cosine similarity* dengan menggunakan persamaan:

$$\textit{Cosine Similarity}\ (d_1, d_2) = \frac{\vec{V}(d1).\vec{V}(d2)}{|\vec{V}(d1)|\,|\vec{V}(d2)|} = \frac{\sum_{i=1}^{t}(\text{tf}-\text{idf}_{(t,d1)}.\text{tf}-\text{idf}_{(t,d2)})}{\sqrt{\sum_{i=1}^{t}\text{tf}-\text{idf}_{(t,d1)}{}^{2}.\sum_{i=1}^{t}\text{tf}-\text{idf}_{(t,d2)}{}^{2}}} \quad (2.5)$$

$|\vec{V}(d1)|$= Panjang Vektor bobot dokumen 1

$|\vec{V}(d2)|$= Panjang Vektor bobot dokumen 2

Contoh metode pembobotan *Frequency-Inverse Document Frequency* dapat diilustrasikan menggunakan 3 kalimat sebagai berikut:

Kalimat 1: Plagiat merupakan perbuatan sengaja atau tidak sengaja dalam pengambilan karangan orang lain atau nilai untuk suatu karya ilmiah

Kalimat 2: Plagiat adalah pengambilan karangan orang lain dan menjadikannya seolah-olah karangan sendiri

Kalimat 3: Plagiat adalah pengambilan karangan orang lain atau nilai untuk suatu karya ilmiah

Langkah 1: Tahap pertama melakukan proses *case folding* pada kalimat 1, kalimat 2 dan kalimat 3

Langkah 2: Melakukan *tokenizing* pada kalimat 1, kalimat 2 dan kalimat 3 untuk mendapatkan *token* atau *term*.

Langkah 3: Melakukan *stopword* pada kata-kata yang tidak memiliki arti penting

Langkah 4: Melakukan *stemming* pada kata-kata yang memiliki imbuhan awal dan imbuhan akhir.

Langkah 5: Kata yang telah didapatkan kemudian dibuat penyusunan daftar kata (kata yang sama hanya akan dimuat satu kali)

Langkah 6: Setiap kata yang ganda hanya ditulis sekali, dan dipisahkan.

Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
© Hak cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau
UIN SUSKA RIAU

Plagiat | ambil | karangan | orang | lain | jadikan | seolah-olah | sendiri | rupa | buat | sengaja | tidak | dalam | nilai | suatu | karya | ilmiah |

Langkah 7: Menghitung bobot tiap kata berdasarkan persamaan yang telah ditetapkan:

Tabel 2.3 Perhitungan bobot token pada kalimat 1

Kata	*Term Frequency* (tf_t)	df_t	idf_t=Log(N/df_t)	$tf_{(t,d)}.idf_t$
Plagiat	1	3	0	0
Rupa	1	1	0,477121255	0,477121255
Buat	1	1	0,477121255	0,477121255
Sengaja	2	1	0,477121255	0,954242509
Dalam	1	1	0,477121255	0,477121255
Ambil	1	3	0	0
Karangan	1	3	0	0
Orang	1	3	0	0
Lain	1	3	0	0
Nilai	1	2	0,176091259	0,176091259
Suatu	1	2	0,176091259	0,176091259
Karya	1	2	0,176091259	0,176091259
Ilmiah	1	2	0,176091259	0,176091259
Jadikan	0	1	0,477121255	0
seolah-olah	0	1	0,477121255	0
Sendiri	0	1	0,477121255	0
Tidak	1	1	0,477121255	0,477121255

Berdasarkan tabel metode pembobotan *term frequency-inverse document frequency* kalimat 1 informasi yang di dapat yaitu pada kalimat 1 ditemukan satu kata yang berulang-ulang yaitu (sengaja), dan ada kesamaan kata antara kalimat 1 dan kalimat 2 yaitu (nilai, suatu, karya,ilmiah) dan terdapat beberapa kesamaan kata dari ketiga kalimat yaitu pada kata (Plagiat, ambil, Karangan, Orang, Lain).

2.4 Kerja Praktek

Kerja Praktek dapat dilakukan di dalam instansi-instansi pemerintah dan perusahaan-perusahaan yang sudah jelas izin usahanya. Kerja praktek merupakan mata kuliah wajib untuk setiap mahasiswa Teknik Informatika, yang mana setiap

Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
© Hak cipta milik UIN Suska Riau State Islamic University of Sultan Syarif Kasim Riau

mahasiswa dapat mengambil mata kuliah kerja praktek jika telah memenuhi beberapa persyaratan. Persyaratan tersebut diantaranya yaitu mahasiswa tersebut telah menyelesaikan mata kuliah minimal sebanyak 80 SKS di semester 6 dan mahasiswa tersebut tidak mengulang mata kuliah wajib seperti Jaringan Komputer, Sistem Informasi, dan Rekayasa Perangkat Lunak. Setiap mahasiswa diminta untuk mencari permasalahan yang ada pada perusahaan atau instansi tersebut serta kemudian mencari solusi dari permasalahan tersebut, yang mana solusinya akan mempermudah pihak instansi dan perusahaan. Resiko kegagalan kerja praktek ini sangat berdampak ketika akan melakukan penelitian Tugas Akhir. Ada beberapa kerugian yang terjadi jika mahasiswa mengalami kegagalan dalam melaksanakan kerja praktek diantaranya:

1. Mahasiswa yang gagal melaksanakan kerja praktek dengan waktu yang telah ditentukan wajib mengulang kembali di periode kerja praktek berikutnya.
2. Setiap mahasiswa yang gagal menyelesaikan kerja prakteknya pada sebuah perusahaan atau instansi, maka pada periode berikutnya mahasiswa tersebut harus memilih perusahaaan atau instansi yang berbeda dari periode sebelumnya.
3. Semakin sering mahasiswa menunda-nunda kerja prakteknya maka kemungkinan untuk melakukan penelitian akan semakin lama pula, sehingga menghambat atau memperlama mahasiswa teknik informatika untuk melakukan penelitian Tugas Akhir.

Mata kuliah kerja praktek merupakan bagian terpenting mahasiswa saat akan melakukan Tugas Akhir atau Skripsi, karena jika mata kuliah kerja praktek ini tidak terselesaikan maka mahasiswa teknik informatika tersebut tidak dapat melanjutkan Tugas Akhirnya, karena kerja praktek merupakan sebuah persyaratan wajib bagi mahasiswa saat akan melakukan penelitian Tugas Akhir. Laporan kerja yang sudah terjilid dan dokumen yang dikirim ke koordinator kerja praktek sebagai bukti akhir bahwa mahasiswa tersebut telah menyelesaikan kerja prakteknya. Ada beberapa hal yang menyebabkan mahasiswa mengalami kegagalan dalam melaksanakan kerja praktek yaitu:

Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
© Hak Cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau
UIN SUSKA RIAU

1. Laporan kerja praktek mahasiswa tersebut teridentifikasi melakukan *plagiarism* terhadap dokumen laporan kerja praktek mahasiswa lainnya.
2. Mahasiswa tersebut tidak menyelesaikan laporan kerja praktek dalam tempo yang telah ditetapkan oleh pihak jurusan teknik informatika.
3. Mahasiswa tersebut tidak sanggup menyelesaikan kerja praktek sesuai dengan permintaan perusahaan atau instansi tersebut.
4. Mahasiswa tersebut lalai saat melakukan bimbingan sehingga saat waktunya hampir habis pelaksanaan bimbingan belum maximal sehingga tidak dapat diseminarkan karena belum mencapai target yang diinginkan. (Sumber: Ketua Jurusan Teknik Informatika)

2.5 *Unified Modeling Language* (*UML*)

Munawar (2005) *Unified Modeling Language* (*UML*) adalah salah satu alat bantu yang sangat handal di dunia pengembangan sistem yang berorientasi objek. *Unified Modeling Language* (*UML*) menyediakan bahasa pemodelan visual yang mudah dimengerti serta dilengkapi dengan mekanisme yang efektif untuk berbagi (*sharing*). *Unified Modeling Language* (*UML*) merupakan suatu kesatuan dari bahassa pemodelan yang dikembangkan oleh *Booch*, *Object Modeling Technique* (*OMT*) dan *Object Oriented Software Engineering* (*OOSE*). *Unified Modeling Language* (*UML*) memiliki banyak diagram yang mendukung terciptanya pemodelan sistem diantaranya yang dapat mewakili bagian terpenting dari suatu sistem, yaitu:

a. *Use Case Diagram*

Use case diagram adalah deskripsi fungsi dari sebuah sistem dari perspektif pengguna, *use case diagram* bekerja dengan mendeskripsikan tipikal interaksi antara pengguna (*user*) sebuah sistem dengan sistemnya sendiri melalui sebuah cerita bagaimana sebuah sistem dipakai dengan *scenario*. Setiap *scenario* mendeskripsikan urutan kejadian.

b. *Sequence Diagram*

Sequence diagram adalah salah satu dari *interaction diagram*. Karena *sequence diagram* mengacu kepada objek, maka sebelum membuat diagram

Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
© Hak cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau

ini *class diagram* sudah teridentifikasi. *UML* memberikan tiga *stereotype* untuk *class yaitu boundary*, *control*, dan *entity*, atau disingkat dengan *BCE*.

c. *Class Diagram*

Class diagram memberikan pandangan secara luas dari suatu sistem dengan menunjukkan kelas-kelasnya dan hubungannya. Objek kelas dapat dianggap sebagai defenisi umum dari satu jenis sistem, memberikan pandangan secara luas dari suatu sistem. *Class diagram* bersifat statis menggambarkan hubungan apa yang terjadi bukan apa yang terjadi jika mereka berhubungan.

d. *Activity Diagram*

Activity diagram adalah teknik untuk mendeskripsikan logika *prosedural*, proses bisnis dan aliran kerja dalam banyak kasus. *Activity diagram* mempunyai peran seperti halnya *flowchart*, akan tetapi perbedaannya dengan *flowchart* adalah *activity diagram* dapat mendukung prilaku *paralel* sedangkan *flowchart* tidak bisa.

2.6 Penelitian Terkait

Penulis memahami beberapa penelitian yang hampir mendekati penelitian yang akan penulis lakukan yaitu "Penentuan Tingkat Kemiripan Laporan Kerja Praktek Menggunakan Metode Pembobotan *Term Frequency-Inverse Document Frequency* dan *Cosine Similarity*". Penelitian yang terkait akan dijelaskan dalam bentuk Tabel 2.4 berikut ini:

Tabel 2.4 Penelitian Terkait

No	Penulis & Tahun	Judul	Metode	Hasil
1	Ajie dkk, 2017	Aplikasi Pendeteksi Dugaan awal plagiarisme pada tugas mahasiswa berdasarkan kemiripan isi teks menggunakan algoritma *Levenshtein Distance*	Algoritma *Levenshtein Distance*	Hasil Pengujian *Similarity* menggunakan ini dapat digunakan tetapi jika struktur kata pada kalimatnya berbeda maka tidak dapat mendeteksi kemiripan.
2	Ryansyah dkk, (2017)	Implementasi Algoritma TF-IDF Pada Pengukuran Kesamaan Dokumen	TF-IDF	Hasil pengujian sistem pengukuran tingkat kesamaan menunjukkan bahwa algoritma TF-IDF membutuhkan paling sedikit 3 dokumen yang digunakan.

Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
© Hak cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau

No	Penulis & Tahun	Judul	Metode	Hasil
3	Mas'udia, dkk, 2017	*Information Retrival* Tugas Akhir dan Perhitungan Kemiripan Dokumen Mengacu pada Abstrak Menggunakan SVM	TF-IDF	Hasil dari penelitian adalah diinputkan kata "Android" maka tampil 4 dokumen yang memiliki kemiripan, dokumen 1 0,9215 dokumen 2 0,5020 dokumen 3 0,2671 dokumen 4 0,1522
4	Siregar dkk, 2017	Aplikasi Penentu Dosen Penguji Skripsi Menggunakan Metode *TF-IDF* dan VSM	TF-IDF	Hasil dari penelitian mendapatkan akurasi 93,22% dalam penentuan dosen penguji dari 59 judul.
5	Amburika dkk, 2016	Teknik VSM Dalam Penentuan Penanganan Dampak Game *Online*	TF-IDF	Hasil pengujian akurasi sistem 80%.
6	Miftahuddin dkk, 2016	Perbandingan *N-GRAM Technique* dan *Rabin Karp* Pada Aplikasi Pendeteksi *Plagiarisme* Dokumen teks Bahasa Indonesia	*N-gram* dan *Rabin Karp*	Hasil dari pendeteksian *similarity* menggunakan metode *N-gram* dan *Rabin Karp*, metode *N-gram* lebih cepat waktu pendeteksian *similarity* dari pada *Rabin Karp*
7	Okfalisa dkk, 2016	Implementasi metode *TF-IDF* dan *Maximum Marginal Relevance Monitoring* Diskusi *Online*	TF-IDF	Hasil penelitian metode TF-IDF berhasil diterapkan untuk monitoring diskusi *online*
8	Sanjaya dkk, 2015	Pengelompokan Dokumen Menggunakan *Winnowing Fingerprint* Dengan Metode *K-Nearest Neighbourd*	*K-Gram* dan *Fingerprint*	Hasil Akurasi dari 10 dokumen persentasi akurasi yang didapat adalah 80%
9	Sugiyamta, 2015	Sistem Deteksi Kemiripan Dokumen Dengan Algoritma *Cosine Similarity* dan *Single Pass Clustering*	Algoritma *Cosine Similarity* dan *Single Pass Clustering*	Hasil penelitian Algoritma *Cosine Similarity* dan *Single Pass Clustering* Dapat dipakai untuk akurasi untuk dokumen abstrak
10	Firdaus dkk, 2014	Aplikasi Pendeteksian kemiripan pada dokumen teks menggunakan Algoritma Nazief & Andriani dan *Cosine Similarity*	Nazief & Adriani *algorithm*	Hasil dari penerapan algoritma Nazief & Adriani dan metode *Cosine Similarity* dapat membandingkan berkas berbeda ekstensi dan membandingkan lebih dari dua dokumen secara bersamaan
11	Salmuasih dkk, 2013	Implementasi Algoritma *Rabin Karp* untuk pendeteksian plagiat Dokumen Teks Menggunakan konsep	*Algorithm Rabin Karp*	Hasil dari penelitian menghasilkan banyaknya konten sebuah file akan memperpanjang waktu prosesnya (*running time*), *stemming* dan *preprocessing* membuat

Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
© Hak cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau

No	Penulis & Tahun	Judul	Metode	Hasil
		Similarity		waktu proses cenderung lebih lama
12	Heriyanto, 2012	Pencarian kemiripan Judul Skripsi dan Abstrak dengan Metode *Exact Match*	*Exact Matching*	Hasil dari pengujian yang dilakukan posisi setiap kata harus sama sehingga persantesi *Similarity* yang didapat kurang konsisten

© Hak cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau
UIN SUSKA RIAU
Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.