

BAB IV

ANALISA DAN PERANCANGAN

4.1 Analisa Sistem Yang Akan Dibangun

Berikut analisa sistem Penentuan Tingkat kemiripan laporan kerja praktek Menggunakan Metode Pembobotan *Term Frequency-Inverse Document Frequency* dan *Cosine Similarity* yang akan dibangun, mengenai Data Masukan (*Input*), Proses, Data Keluaran (*Output*) yaitu:

4.1.1 Data Masukan (*Input*)

Data masukan (*input*) yang diperlukan untuk pengujian kemiripan dokumen kerja praktek adalah:

1. Dokumen Sumber

Dokumen sumber terdiri dari beberapa dokumen kerja praktek yang dijadikan dokumen sumber berdasarkan data laporan kerja praktek semester ganjil 2016/2017 yang telah diseminarkan.

2. Dokumen Uji

Dokumen uji terdiri dari satu laporan kerja praktek yang terindikasi melakukan *plagiarism*, diambil dari salah satu dokumen kerja praktek berdasarkan data laporan kerja praktek semester ganjil 2016/2017 yang telah diseminarkan.

4.1.2 Proses

Proses-proses yang dilakukan pada sistem Penentuan Tingkat kemiripan laporan kerja praktek Menggunakan Metode Pembobotan *Term Frequency-Inverse Document Frequency* dan *Cosine Similarity* diantaranya:

1. Pencarian Kemiripan

Proses pencarian kemiripan dokumen kerja praktek antara beberapa dokumen sumber dengan satu dokumen uji yang menghasilkan jumlah persentasi kemiripan dan rincian perhitungan metode pembobotan TF-IDF.

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.

4.1.3 Data Keluaran (*Output*)

Data keluaran (*Output*) yang akan dikeluarkan sistem Penentuan Tingkat kemiripan laporan kerja praktek menggunakan Metode Pembobotan *Term Frequency-Inverse Document Frequency* dan *Cosine Similarity* diantaranya:

1. Persentasi Kemiripan

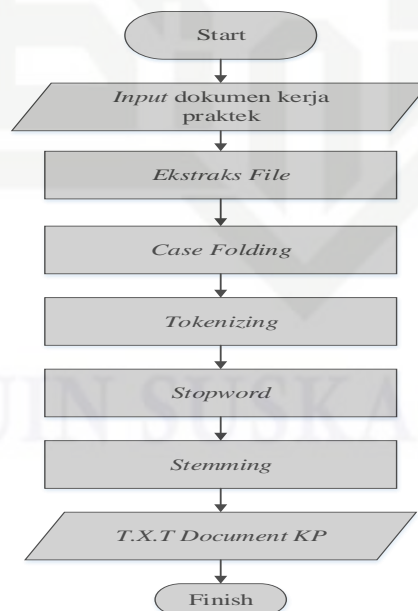
Persentasi kemiripan didapat dari perbandingan antara dokumen uji dengan dokumen sumber menggunakan metode pembobotan *term frequency-inverse document frequency* dan *Cosine Similarity*.

2. Rincian Laporan

Rincian laporan berisi perhitungan persentasi kemiripan dokumen kerja praktek secara terperinci.

4.2 Analisa Pemrosesan Teks

Sebelum dilakukan pendeteksian kemiripan dokumen pada dokumen KP maka harus melakukan pemrosesan teks dengan melalui beberapa proses tahapan-tahapan agar dapat dilakukan untuk tahapan selanjutnya, tahapan-tahapannya dapat dilihat pada *flowchart* berikut ini:

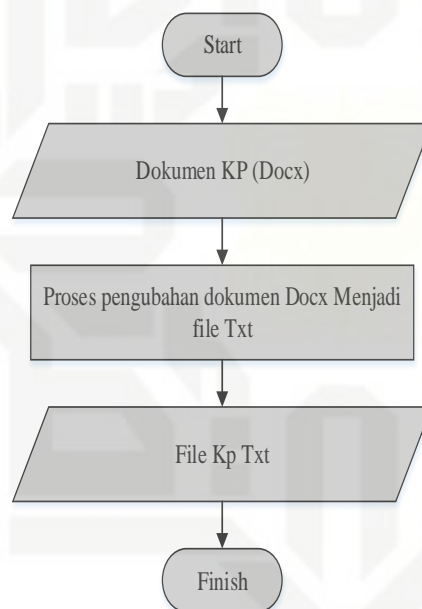


Gambar 4.1 Flowchart Pemrosesan Teks Dokumen KP

Berdasarkan *flowchart* diatas dapat dijelaskan tahapan-tahapan dalam pembuatan kata kunci pada dokumen kerja praktek. Adapun penjelasan dari *flowchart* diatas dapat dilihat di bawah ini:

4.2.1 Ekstraks File

Ekstraks file adalah proses pengubahan format dokumen laporan kerja praktek yang berformat docx ke dalam format txt untuk menghilangkan file gambar dan tabel, karena yang diproses untuk menentukan tingkat kemiripan dokumen laporan kerja praktek hanya teks yang ada di dalam dokumen kerja praktek, penjelasan proses *ekstraks file* dapat dilihat pada *flowchart* dibawah ini:



Gambar 4.2 Flowchart Ekstraks File

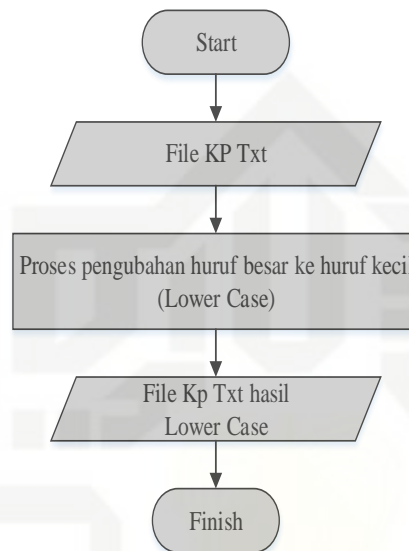
Berikut ini merupakan *pseudocode* proses *ekstraks file* dokumen kerja praktek sebelum dilakukan pemrosesan teks:

```

99 //proses ekstraks File
100 public function convertToText() {
101
102     if(isset($this->filename) && !file_exists($this->filename)) {
103         return "File Not exists";
104     }
105
106     $fileArray = pathinfo($this->filename);
107     $file_ext = $fileArray['extension'];
108     if($file_ext == "doc" || $file_ext == "docx" || $file_ext == "xlsx" || $file_ext == "pptx")
109     {
110         if($file_ext == "doc") {
111             return $this->read_doc();
112         } elseif($file_ext == "docx") {
113             return $this->read_docx();
114         } elseif($file_ext == "xlsx") {
115             return $this->xlsx_to_text();
116         } elseif($file_ext == "pptx") {
117             return $this->pptx_to_text();
118         }
119     } else {
120         return "Invalid File Type";
121     }
122 }
  
```

4.2.2 Case Folding

Case folding merupakan proses pengubahan seluruh kata yang mengandung huruf besar diubah menjadi huruf kecil. Penjelasan proses *Case Folding* dapat dilihat pada *flowchart* dibawah ini:



Gambar 4.3 Flowchart Case Folding

Berikut ini merupakan *peudocode* proses *Case folding* dokumen kerja praktek sebelum dilakukan pemrosesan teks:

```
$token = preg_split("/[\\d\\W\\s]+/", strtolower($load_file));
```

Proses *case folding* dapat dilihat pada Tabel 4.1 berikut:

Tabel 4.1 Proses Case Folding

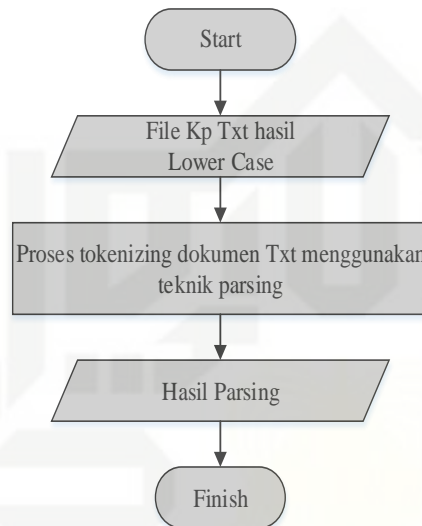
NO	Dokumen	Case Folding
DOK 1	Plagiat merupakan perbuatan sengaja atau tidak sengaja dalam pengambilan karangan orang lain atau nilai untuk suatu karya ilmiah.	plagiat merupakan perbuatan sengaja atau tidak sengaja dalam pengambilan karangan orang lain atau nilai untuk suatu karya ilmiah
DOK 2	Plagiat adalah pengambilan karangan orang lain dan menjadikannya seolah-olah karangan sendiri	plagiat adalah pengambilan karangan orang lain dan menjadikannya seolah-olah karangan sendiri
DOK 3	Plagiat adalah pengambilan karangan orang lain atau nilai untuk suatu karya ilmiah.	plagiat adalah pengambilan karangan orang lain atau nilai untuk suatu karya ilmiah

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.

4.2.3 Tokenizing

Tokenizing merupakan proses pemotongan teks berdasarkan hasil proses *case folding* menjadi perkata dan disusun berurutan kembali berdasarkan kalimat aslinya, untuk kata yang berulang hanya ditulis sekali saja, Penjelasan proses *Case Folding* dapat dilihat pada *flowchart* dibawah ini:



Gambar 4.4 Flowchart Tokenizing

Berikut ini merupakan *pseudocode* proses *Tokenizing* dokumen kerja praktek dari teks dokumen kerja praktek:

```

19 $token = preg_split("/[\d\W\s]+/", strtolower($load_file));
20 if ($parsing==1){
21   $token=$token;
22 }else if ($parsing==2){
23   foreach ($token as $key => $value) {
24     if (count($token)>($key+1)){
25       $v[]=$value." ".$token[$key+1].",";
26     }
27   }
28   $vn=implode(" ", $v);
29   $token=explode(" ", $vn);
30
31   foreach ($token as $key => $value) {
32     if ($value){
33       $tokenvalue[]=$value;
34     }
35   }
36
37   $token=$tokenvalue;
38 }else if ($parsing==3){
39   foreach ($token as $key => $value) {
40     if (count($token)>($key+1)){
41       $v[]=$value." ".$token[$key+1]." ".$token[$key+2].",";
42     }
43   }
44   $vn=implode(" ", $v);
45   $token=explode(" ", $vn);
46
47   foreach ($token as $key => $value) {
48     if ($value){
49       $tokenvalue[]=$value;
50     }
51   }
  
```

- 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.

```

63 $token=$tokenvalue;
64 }else if ($parsing==4){
65     foreach ($token as $key => $value) {
66         if (count($token)>($key+1)){
67             $v[]=$value." ".$token[$key+1]." ".$token[$key+2]." ".$token[$key+3].",";
68         }
69     }
70     $vn=implode(" ", $v);
71     $token=explode(" ", $vn);
72
73     foreach ($token as $key => $value) {
74         if ($value){
75             $tokenvalue[]=$value;
76         }
77     }
78
79     $token=$tokenvalue;
80 }

```

Proses *tokenizing* dapat dilihat pada Tabel 4.2 berikut ini:

Tabel 4.2 Proses *Tokenizing*

NO	Case Folding	Tokenizing
DOK 1	plagiat merupakan perbuatan sengaja atau tidak sengaja dalam pengambilan karangan orang lain atau nilai untuk suatu karya ilmiah	plagiat merupakan perbuatan tidak sengaja dalam pengambilan karangan orang lain atau nilai untuk suatu karya ilmiah
DOK 2	plagiat adalah pengambilan karangan orang lain dan menjadikannya seolah-olah karangan sendiri	plagiat adalah pengambilan karangan orang lain dan menjadikannya seolah-olah sendiri
DOK 3	plagiat adalah pengambilan karangan orang lain atau nilai untuk suatu karya ilmiah	plagiat adalah pengambilan karangan orang lain atau nilai untuk suatu karya ilmiah

4.2.4 *Stopword*

Proses *stopword* dilakukan setelah proses *tokenizing* selesai, *stopword* merupakan proses pengolahan teks dengan menghapus kata-kata yang tidak relevan atau tidak memiliki arti seperti kata hubung, namun pada penelitian ini ada yang menggunakan proses *stopword* dan tanpa proses *stopword* untuk melihat mana lebih akurat hasilnya, Penjelasan proses *stopword* dapat dilihat pada *flowchart* dibawah ini:

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.



Gambar 4.5 Flowchart Stopword

Berikut ini merupakan *pseudocode* proses *stopword* dokumen kerja praktek dari teks dokumen kerja praktek:

```

9      $stopwords = file_get_contents("../mod/stopwords.txt");
10     $stopwords = preg_split("/[\s]+/", $stopwords);
  
```

Proses *stopword* dapat dilihat pada Tabel 4.3 berikut ini:

Tabel 4.3 Proses Stopword

NO	Tokenizing	Stopword	Hasil Stopword
DOK 1	plagiat merupakan perbuatan tidak sengaja dalam pengambilan karangan orang lain atau nilai untuk suatu karya ilmiah	Atau	plagiat merupakan perbuatan tidak sengaja dalam pengambilan karangan orang lain nilai untuk suatu karya ilmiah
DOK 2	plagiat adalah pengambilan karangan orang lain dan menjadikannya seolah-olah sendiri	adalah dan	plagiat pengambilan karangan orang lain menjadikannya seolah-olah sendiri
DOK 3	plagiat adalah pengambilan karangan orang lain atau nilai untuk suatu karya ilmiah	Adalah atau	plagiat pengambilan karangan orang lain nilai untuk suatu karya ilmiah

4.2.5 Stemming

Stemming proses pengolahan teks tahap terakhir setelah proses penghilangan kata-kata yang tidak relevan (*stopword*), proses *stemming* yaitu proses penghilangan imbuhan pada kata, proses *stemming* dapat dilihat pada *flowchart* dibawah ini:

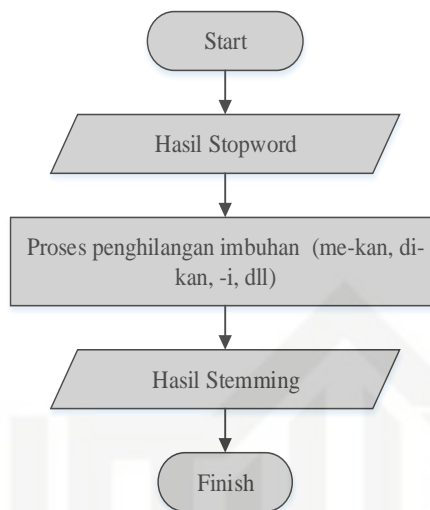
Hak Cipta Dilindungi Undang-Undang

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.



Gambar 4.6 Flowchart Stemming

Berikut ini merupakan *pseudocode* proses *Stemming* dokumen kerja praktek dari teks dokumen kerja praktek:

```

72  if($stemming==1){
73      $token = array_diff($token, $stopwords);
74  }
  
```

Proses *stemming* pada imbuhan kata dapat dilihat pada Tabel 4.4 yaitu:

Tabel 4.4 Proses *stemming*

NO	Hasil Stopword	Stemming	Hasil Stemming
DOK 1	plagiat merupakan perbuatan tidak sengaja dalam pengambilan karangan orang lain nilai untuk suatu karya ilmiah	Me-....-kan Per-....-an Peng-....-an -an	plagiat rupa buat tidak sengaja dalam ambil karang orang lain nilai untuk suatu karya ilmiah
DOK 2	plagiat pengambilan karangan orang lain menjadikannya seolah-olah sendiri	Peng-....-an -an Men-....-nya	plagiat ambil karang orang lain jadikan seolah-olah sendiri
DOK 3	plagiat pengambilan karangan orang lain nilai untuk suatu karya ilmiah	Peng-....-an -an	plagiat ambil karang orang lain nilai untuk suatu karya ilmiah

4.3 Metode Pembobotan *Term Frequency-Inverse document Frequency*

Setelah dilakukan pemrosesan teks pada dokumen kerja praktek, selanjutnya menerapkan metode pembobotan *term frequency-inverse document frequency*

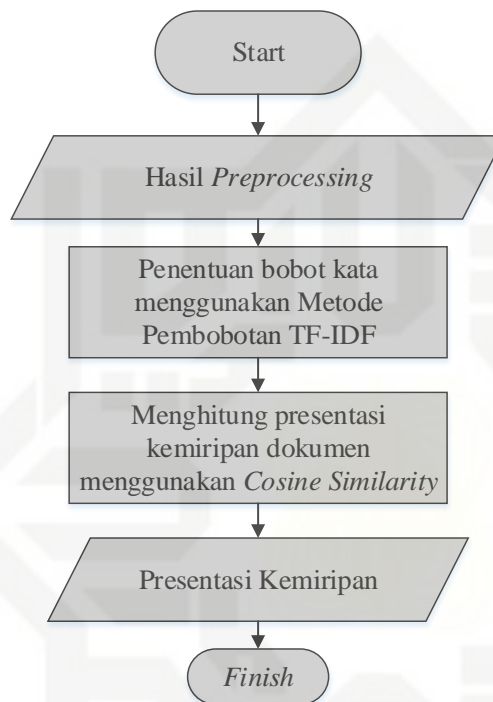
Hak Cipta Dilindungi Undang-Undang

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.

untuk melakukan pembobotan *term* pada kata kunci yang telah dihasilkan dari pemrosesan teks dan menghitung tingkat kemiripan dokumen menggunakan *cosine similarity*, adapun tahapan-tahapan prosesnya dapat dilihat pada *flowchart* dibawah ini:



Gambar 4.7 Flowchart Metode Pembobotan TF-IDF

4.3.1 Input Dokumen Kerja Praktek

Input dokumen kerja praktek ini merupakan data dokumen teks laporan kerja praktek yang telah melakukan pemrosesan *ekstraks file*, *case folding*, *tokenizing*, *stopword*, dan *stemming*. Setelah didapat hasil pemrosesan teks tahap selanjutnya yaitu dilakukan perhitungan bobot *term* (kata) pada setiap dokumen kerja praktek dengan menerapkan konsep metode pembobotan *term frequency-inverse document frequency* dan *Cosine Similarity*.

4.3.2 Metode Pembobotan *Term Frequency-Inverse Document Frequency*

Berdasarkan hasil pemrosesan teks yang dihasilkan selanjutnya dilakukan perhitungan bobot *term* dengan menghitung *term frequency* menggunakan persamaan (2.1), *inverse document frequency* menggunakan persamaan (2.2), dan menghitung *term frequency-inverse document frequency* menggunakan persamaan

(2.3). Pada kasus pembobotan ini penulis menggunakan teks lengkap tanpa proses *stopword* dan *stemming* untuk melihat tingkat kemiripan suatu dokumen. Pada contoh perhitungan bobot dokumen ini penulis menggunakan 3 kalimat yang dapat dilihat dibawah ini:

kalimat 1 : Plagiat adalah pengambilan karangan orang lain atau nilai untuk suatu karya ilmiah.

Kalimat 2 : Plagiat merupakan perbuatan sengaja atau tidak sengaja dalam pengambilan karangan orang lain atau nilai untuk suatu karya ilmiah.

kalimat 3 : Plagiat adalah pengambilan karangan orang lain dan menjadikannya seolah-olah karangan sendiri

Tabel 4.5 pembobotan kata

Kata	Term Frequency			df _t	idf _t	tf-idf _(t,d)			Bobot Dokumen		Panjang Vektor		
	tf _(t,d1)	Tf _(t,d2)	tf _(t,d3)			tf-idf _(t,d1)	tf-idf _(t,d2)	tf-idf _(t,d3)	$\vec{V}(d1) \cdot \vec{V}(d2)$	$\vec{V}(d1) \cdot \vec{V}(d3)$	$ \vec{V}(d1) $	$ \vec{V}(d2) $	$ \vec{V}(d3) $
plagiat	1	1	1	3	0	0	0	0	0	0	0	0	0
adalah	1	0	1	2	0,176	0,176	0	0,176	0	0,031	0,031	0	0,031
pengambilan	1	1	1	3	0	0	0	0	0	0	0	0	0
karangan	1	1	2	3	0	0	0	0	0	0	0	0	0
orang	1	1	1	3	0	0	0	0	0	0	0	0	0
lain	1	1	1	3	0	0	0	0	0	0	0	0	0
atau	1	2	0	2	0,176	0,176	0,352	0	0,062	0	0,031	0,124	0
nilai	1	1	0	2	0,176	0,176	0,176	0	0,031	0	0,031	0,031	0
untuk	1	1	0	2	0,176	0,176	0,176	0	0,031	0	0,031	0,031	0
suatu	1	1	0	2	0,176	0,176	0,176	0	0,031	0	0,031	0,031	0
karya	1	1	0	2	0,176	0,176	0,176	0	0,031	0	0,031	0,031	0
ilmiah	1	1	0	2	0,176	0,176	0,176	0	0,031	0	0,031	0,031	0
dan	0	0	1	1	0,477	0	0	0,477	0	0	0	0	0,2276
menjadikannya	0	0	1	1	0,477	0	0	0,477	0	0	0	0	0,2276
seolah-olah	0	0	1	1	0,477	0	0	0,477	0	0	0	0	0,2276
sendiri	0	0	1	1	0,477	0	0	0,477	0	0	0	0	0,2276
merupakan	0	1	0	1	0,477	0	0,477	0	0	0	0	0,228	0
perbuatan	0	1	0	1	0,477	0	0,477	0	0	0	0	0,228	0
sengaja	0	2	0	1	0,477	0	0,954	0	0	0	0	0,911	0
tidak	0	1	0	1	0,477	0	0,477	0	0	0	0	0,228	0
dalam	0	1	0	1	0,477	0	0,477	0	0	0	0	0,228	0

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.

4.3.3 Perhitungan Presentasi Kemiripan

Setelah dilakukan perhitungan bobot *term* selanjutnya dihitung presentasi kemiripan antara dokumen uji dengan dokumen lainnya, pada penelitian ini penulis menggunakan 3 kalimat contoh. kalimat 2 dan kalimat 3 sebagai dokumen sumber, kalimat 1 sebagai dokumen uji. Perhitungan kemiripan suatu dokumen menggunakan persamaan *cosine similarity* (2.5). Contoh perhitungannya dapat dilakukan berdasarkan tahapan-tahapan dibawah ini:

1. Berdasarkan pembobotan kata pada Tabel 4.5 tentukan nilai bobot kata dokumen uji dan dokumen sumber dengan mengalikan nilai TF-IDF dokumen uji dan nilai TF-IDF dokumen sumber menggunakan persamaan (2.3) hasilnya dapat dilihat seperti dibawah ini:

Bobot dokumen uji dan dokumen sumber 1 $\vec{V}(d1) \cdot \vec{V}(d2) = (0 * 0) + (0,176091 * 0) + (0 * 0) + (0 * 0) + (0 * 0) + (0,176091 * 0,352183) + (0,176091 * 0,176091) + (0,176091 * 0,176091) + (0,176091 * 0,176091) + (0,176091 * 0,176091) + (0,176091 * 0,176091) + (0,176091 * 0,176091) + (0 * 0,477121) + (0 * 0,477121) + (0 * 0,954243) + (0 * 0,477121) = 0,217057$

Bobot dokumen uji dan dokumen sumber 2 $\vec{V}(d1) \cdot \vec{V}(d3) = (0 * 0) + (0,176091 * 0,176091) + (0 * 0) + (0 * 0) + (0 * 0) + (0 * 0) + (0,176091 * 0) + (0,176091 * 0) + (0,176091 * 0) + (0,176091 * 0) + (0,176091 * 0) + (0,176091 * 0) + (0 * 0,477121) + (0 * 0,477121) + (0 * 0,477121) + (0 * 0,477121) = 0,031008$

2. Menentukan panjang vektor kata dokumen sumber dengan mengangkat duakan nilai TF-IDF menggunakan persamaan (2.4), berdasarkan pembobotan Dokumen Uji pada Tabel 4.5, yang mana nilai Dokumen Uji merupakan pembobotan kata kunci sama yang dimiliki antara dokumen uji dan dokumen sumber seperti dibawah ini:

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- 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.

$$\text{Panjang Vektor } |\vec{V}(d1)| = \sqrt{\begin{array}{l} (0^2) + (0,176091^2) + (0^2) \\ + (0^2) + (0^2) + (0^2) \\ + (0,176091^2) \\ + (0,176091^2) + (0,176091^2) \\ + (0,176091^2) + (0,176091^2) \\ + (0,176091^2) + (0^2) \\ + (0^2) + (0^2) \\ + (0^2) + (0^2) \end{array}} = 0,465894$$

- Menentukan panjang vektor kata kalimat 2 dengan menjumlahkan nilai TF-IDF menggunakan persamaan (2.4), berdasarkan pembobotan kalimat 2 pada Tabel 4.5. Sebelum dilakukan perhitungan panjang vektor kalimat 2, maka terlebih dahulu dihitung bobot kata kalimat 2 seperti dibawah ini:

$$\text{Panjang Vektor } |\vec{V}(d2)| = \sqrt{\begin{array}{l} (0^2) + (0^2) + (0^2) \\ + (0^2) + (0^2) + (0^2) \\ + (0,352183^2) + (0,176091^2) \\ + (0,176091^2) + (0,176091^2) \\ + (0,176091^2) + (0,176091^2) \\ + (0,477121^2) + (0,477121^2) \\ + (0,954243^2) + (0,477121^2) \\ + (0,477121^2) \end{array}} = 1,449217$$

- Menentukan panjang vektor kalimat 3 dengan menjumlahkan nilai TF-IDF menggunakan persamaan (2.4), berdasarkan pembobotan kalimat 3 pada Tabel 4.5. Sebelum dilakukan perhitungan panjang vektor kalimat 3, maka terlebih dahulu dihitung bobot kata kalimat 3 seperti dibawah ini:

$$\text{Vektor bobot kata } |\vec{V}(d3)| = \sqrt{\begin{array}{l} (0^2) + (0,176091^2) + (0^2) \\ + (0^2) + (0^2) + (0^2) + (0^2) \\ + (0^2) + (0^2) + (0^2) + (0^2) \\ + (0^2) + (0,477121^2) \\ + (0,477121^2) + (0,477121^2) \\ + (0,477121^2) \end{array}} = 0,970354$$

- Setelah didapat jumlah bobot kata dokumen sumber dan dokumen uji, serta panjang vektor kata kunci setiap dokumen lalu dihitung persentasi kemiripannya menggunakan *cosine similarity* menggunakan persamaan (2.5),

untuk kemiripan kalimat 1, kalimat 2 dan kalimat 3 dapat dilihat pada perhitungan dibawah ini:

$$\text{Cosine Similarity } (d_1, d_2) = \frac{[(0,217057)]}{[(0,465894) \times (1,449217)]} = 0,321 * 100\% = 32,1\%$$

$$\text{Cosine Similarity } (d_1, d_3) = \frac{[(0,031008)]}{[(0,465894) \times (0,970354)]} = 0,069 * 100\% = 6,9\%$$

4.3.4 Hasil Presentasi Kemiripan

Jadi setelah dilakukan proses pembobotan *term* (kata) dokumen uji dan dokumen sumber dicari nilai kata kunci dan panjang vektor kata kuncinya, maka dapat disimpulkan presentasi kemiripan dokumen berdasarkan *cosine similarity* didapat kemiripan kalimat 1 (d1), dengan kalimat 2 (d2) dan kalimat 3 (d3) yaitu 32,1% dan 6,9%.

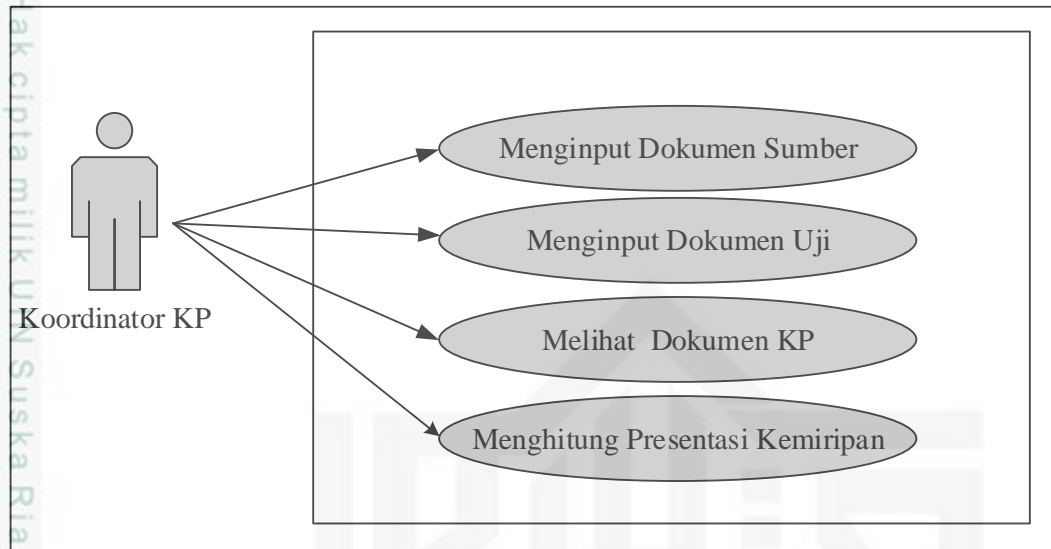
4.4 Perancangan Sistem

Pada tahapan ini dilakukan perancangan sistem berdasarkan analisa yang dilakukan sebelumnya, adapun perancangannya menggunakan *Use Case Diagram*, Perancangan struktur menu dan perancangan *interface*.

4.4.1 Use Case Diagram

Pada *Use Case Diagram* menjelaskan siapa saja aktor yang terlibat di dalam sistem dan menjelaskan apa saja proses-proses yang akan dilakukan terhadap sistem. Aktor merupakan pengguna yang berhubungan secara langsung dengan sistem, yang dapat menginput dokumen dan menerima informasi dari sistem. *Use Case Diagram* yang akan dibangun dapat dilihat pada Gambar 4.8:

UIN SUSKA RIAU



Gambar 4.8 Use case diagram

Berdasarkan Gambar 4.8 memiliki satu aktor yaitu koordinator kerja praktek. Koordinator kerja praktek dapat menginputkan data dokumen kerja praktek yang memiliki kemiripan antar dokumen, melihat data dokumen KP yang akan dijadikan dokumen sumber dan dapat melihat presentasi kemiripan dokumen uji.

4.4.1 Use Case Spesification

Use Case Spesification berisi penjelasan dari setiap *Use Case Diagram* yang ada pada Gambar 4.8 yaitu:

Tabel 4.6 Use Case Spesification

No	Use Case	Penjelasan
1	Menginputkan Dokumen Sumber	<i>Use Case</i> ini berisikan proses penginputan dokumen kerja praktek yang akan dijadikan sebagai dokumen sumber, penginputan dilakukan oleh koordinator kerja praktek .
2	Menginputkan Dokumen Sumber	<i>Use Case</i> ini berisikan proses penginputan dokumen uji kerja praktek yang terindikasi melakukan plagiat dengan dokumen sumber.

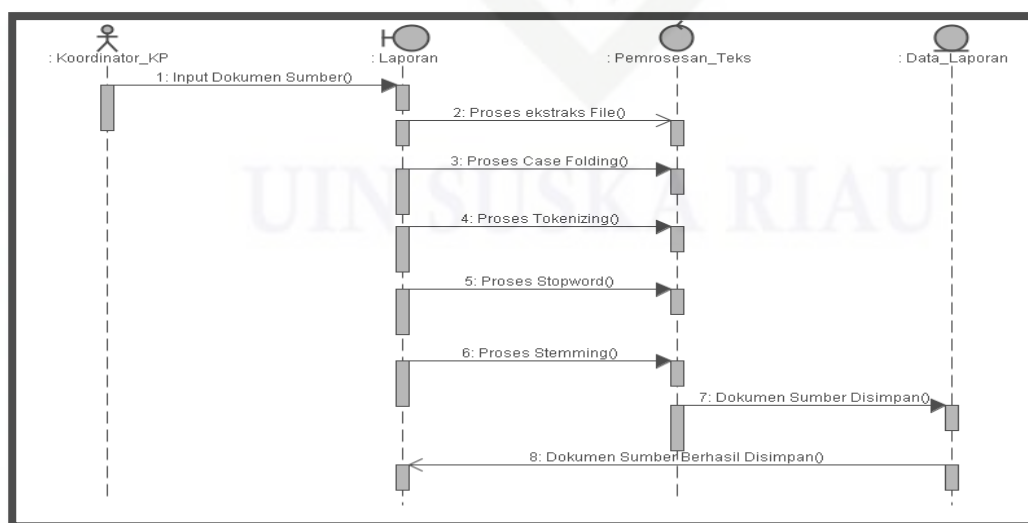
No	Use Case	Penjelasan
2	Melihat dokumen kerja praktek	<i>Use Case</i> ini berisi kumpulan dokumen laporan kerja praktek mahasiswa Teknik Informatika yang akan dijadikan pembanding dengan dokumen uji yang terindikasi melakukan plagiat.
3	Menghitung Presentasi Kemiripan	<i>Use Case</i> ini berisi proses pendeteksian kemiripan antar dokumen uji yang dicurigai memiliki kesamaan dengan beberapa dokumen sumber.

4.4.2 Sequence Diagram

Sequence diagram merupakan proses alur dari sistem yang akan dibuat, penggambaran proses Penentuan Tingkat kemiripan laporan kerja praktek Menggunakan Metode Pembobotan *Term Frequency-Inverse Document Frequency* dan *Cosine Similarity* dapat dilihat pada *sequence diagram* berikut ini:

1. Sequence Diagram Laporan

Sequence diagram laporan merupakan proses penggambaran alur pemrosesan teks dokumen kerja praktek yang akan di simpan dan akan dijadikan sebagai dokumen sumber, prosesnya dapat dilihat pada *sequence diagram* berikut ini:



Gambar 4.9 Sequence Diagram Laporan

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

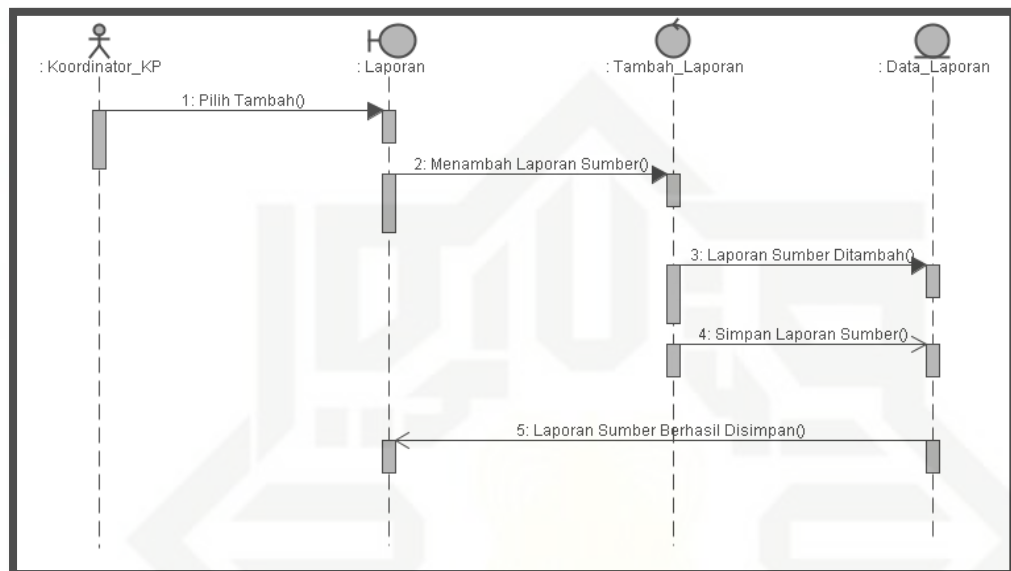
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

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.

a. Sequence Diagram Tambah Laporan

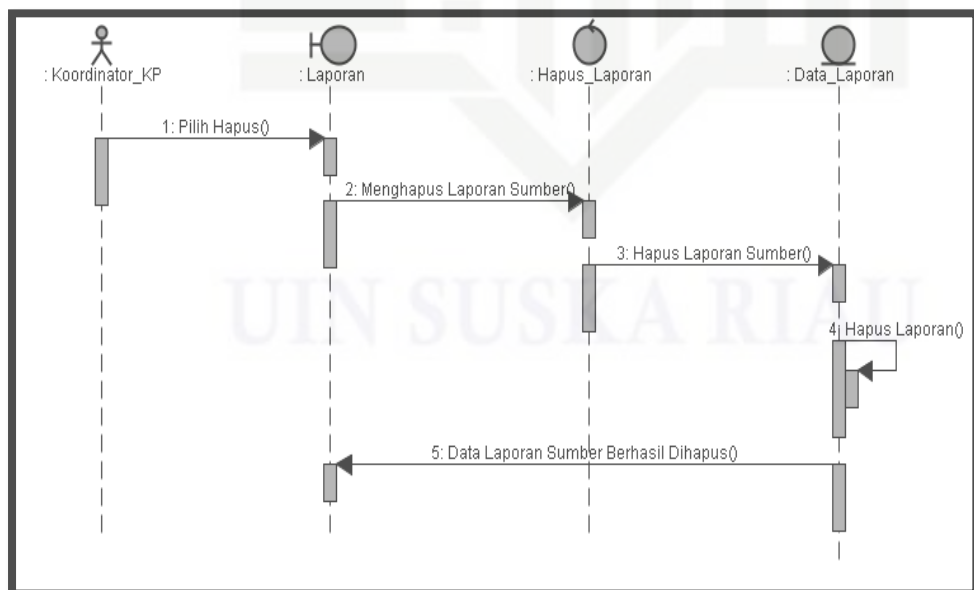
Sequence diagram tambah laporan merupakan proses penambahan dokumen kerja praktek, prosesnya dapat dilihat pada *sequence diagram* berikut ini:



Gambar 4.10 Sequence Diagram Tambah Laporan

b. Sequence Diagram Hapus Laporan

Dibawah ini merupakan penjelasan mengenai *sequence diagram* untuk menghapus dokumen sumber:



Gambar 4.11 Sequence Diagram Hapus Laporan

Hak Cipta Dilindungi Undang-Undang

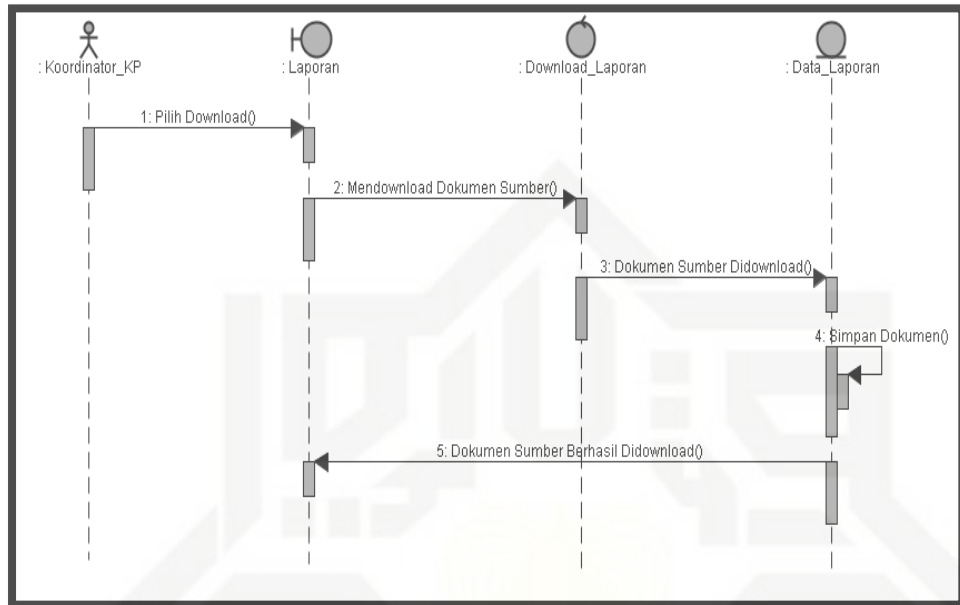
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- 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.

c. Sequence Diagram Download Laporan

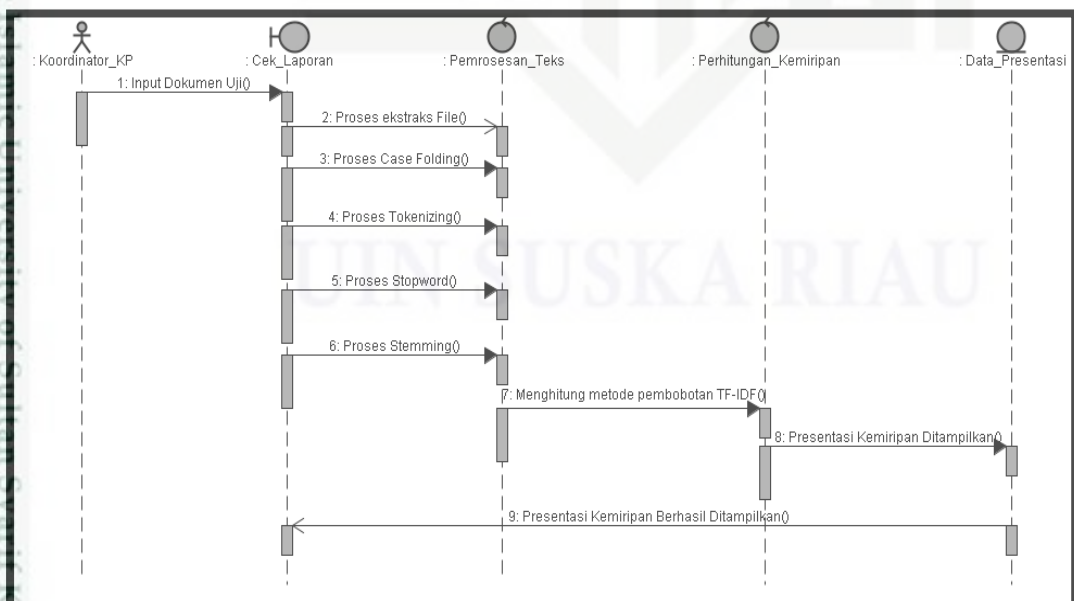
Dibawah ini merupakan penjelasan mengenai *sequence diagram* untuk mendownload dokumen sumber:



Gambar 4.12 Sequence Diagram Download Laporan

2. Sequence Diagram Cek Laporan

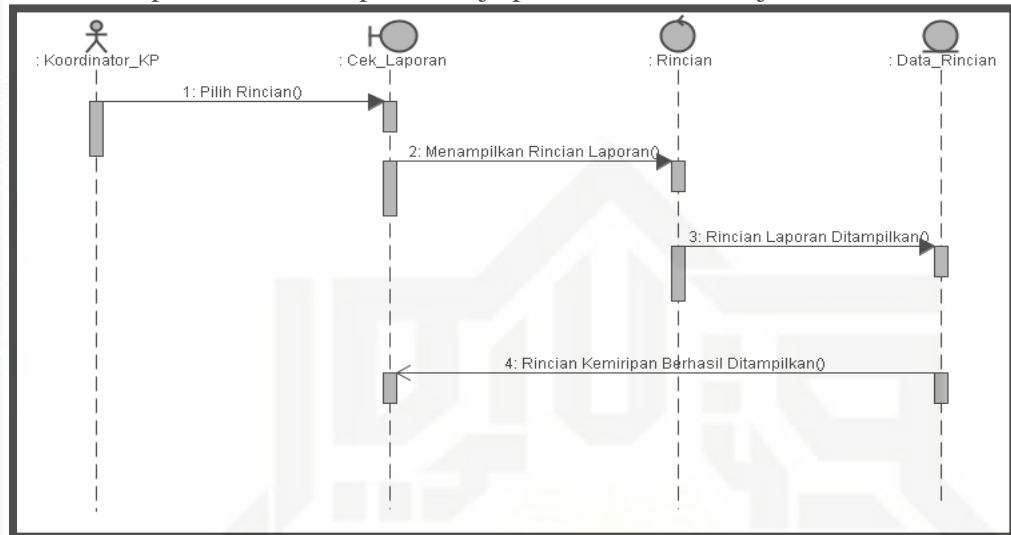
Sequence diagram cek laporan menjelaskan proses pengecekan kemiripan dokumen laporan kerja praktek dengan menggunakan metode pembobotan TF-IDF, prosesnya dapat dilihat pada *sequence diagram* berikut ini:



Gambar 4.13 Sequence Diagram Cek Laporan

a. *Sequence Diagram* Rincian Kemiripan Dokumen

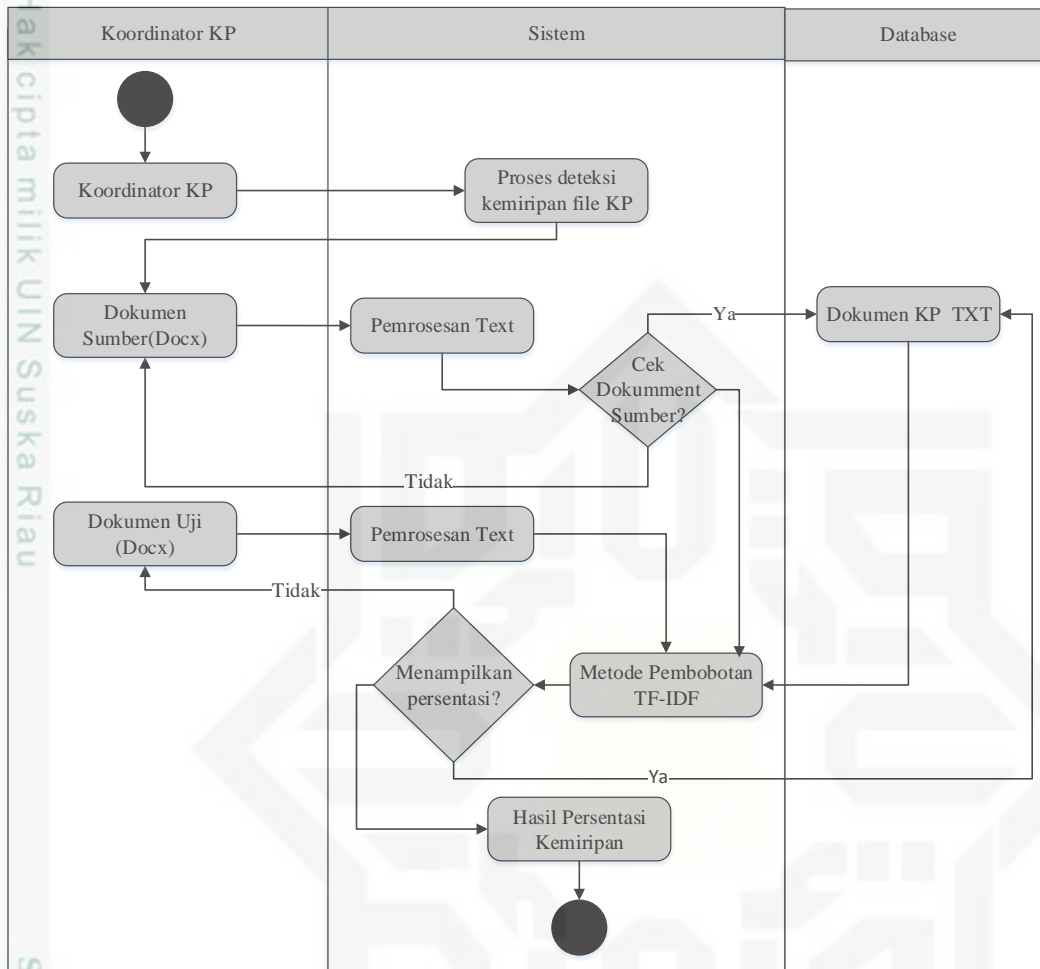
Sequence diagram rincian kemiripan dokumen menjelaskan proses menampilkan rincian laporan kerja praktek dokumen uji:



Gambar 4.14 *Sequence Diagram* Rincian Kemiripan Dokumen

4.4.3 *Activity Diagram*

Activity diagram menggambarkan aktivitas yang dilakukan di dalam sistem yang sedang dirancang bagaimana masing-masing aliran urutan aktifitas pada sistem. *Activity Diagram* dapat dibagi menjadi beberapa jalur kelompok yang menunjukkan obyek yang mana yang bertanggung jawab untuk suatu aktifitas. Peralihan tunggal (*single transition*) timbul dari setiap adanya *activity* (aktifitas) yang saling menghubungi pada aktifitas berikutnya.



Gambar 4.15 Activity diagram

Pada Gambar 4.15 menjelaskan mengenai alur sistem yang akan dibuat yaitu sistem hanya memiliki satu aktor (Koordinator KP), sistem memiliki dua menu yaitu laporan (dokumen sumber) dan cek laporan (dokumen uji). Proses Gambar 4.15 dapat dijelaskan sebagai berikut:

1. Sistem menampilkan menu laporan dan cek laporan.
2. Koordinator kerja praktek yang akan melakukan pengecekan dokumen yang terindikasi melakukan *plagiarism* menginputkan dokumen sumber. Dokumen sumber tersebut harus melalui pemrosesan teks (*ekstraks file, case folding, stopword, stemming*), selanjutnya dokumen sumber dilakukan pengecekan apakah dokumen sumber tersebut layak dijadikan dokumen sumber atau tidak. Dokumen sumber yang tidak layak adalah dokumen yang terindikasi

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.

plagiat ketika dilakukan pemrosesan teks sehingga tidak dapat dijadikan dokumen sumber, dan dokumen yang layak adalah dokumen sumber yang tidak terindikasi melakukan plagiat disimpan ke dalam *database* sistem.

3. Koordinator kerja praktek menginputkan dokumen uji, dokumen uji harus melalui pemrosesan teks (*ekstraks file, case folding, stopword, stemming*), selanjutnya dokumen uji dibandingkan tingkat kemiripannya dengan dokumen sumber yang ada di *database* sistem menggunakan metode pembobotan *term frequency-inverse document frequency*. Setelah dilakukan perhitungan kemiripan dokumen menggunakan metode pembobotan *term frequency-inverse document frequency* sistem menampilkan persentasi kemiripan, ketika persentasi kemiripan ditampilkan maka dokumen uji tersebut layak disimpan atau tidak ke dalam *database* sistem berdasarkan persentasi kemiripan yang telah tampilkan sistem.


4.5 Perancangan *interface*

Perancangan *interface* atau antar muka merupakan tampilan yang akan digunakan dalam Penentuan Tingkat kemiripan laporan kerja praktek Menggunakan Metode Pembobotan *Term Frequency-Inverse Document Frequency* dan *Cosine Similarity*. Pada perancangan *interface* sistem hanya memiliki dua menu utama yang memiliki fungsi berbeda-beda yaitu menu cek laporan dan menu data laporan.

4.5.1 Halaman Menu Laporan

Pada halaman menu laporan berisi data sumber dokumen kerja praktek yang akan dijadikan sebagai dokumen pembanding dengan dokumen uji yang terindikasi melakukan *plagiarism*, rancangan tampilan menu laporan 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.



**PENENTUAN TINGKAT KEMIRIPAN LAPORAN KERJA PRAKTEK
MENGUNAKAN METODE PEMBOBOTAN TERM FREQUENCY-INVERSE
DOCUMENT FREQUENCY DAN COSINE SIMILARITY**

Cek Laporan
Dokumen Laporan

☐ 1 kata
☐ 2 kata
☒ 3 kata
☐ 4 kata

☐ Stemming dan Stopword


Tambah

NO	NIM	NAMA	JUDUL	Tingkat Kemiripan
				Download Delete

Gambar 4.16 Tampilan Menu Laporan

4.5.2 Halaman Tambah Laporan

Halaman tambah laporan merupakan menu untuk menambahkan dokumen sumber yang memiliki 4 item. Tampilan menu tambah laporan dapat dilihat dibawah ini:



**PENENTUAN TINGKAT KEMIRIPAN LAPORAN KERJA PRAKTEK
MENGUNAKAN METODE PEMBOBOTAN TERM FREQUENCY-INVERSE
DOCUMENT FREQUENCY DAN COSINE SIMILARITY**

Cek Laporan
Dokumen Laporan

Simpan


Gambar 4.17 Tampilan Menu Tambah Laporan

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.

4.5.3 Halaman Cek Laporan

Halaman cek laporan merupakan halaman *penginputan* dokumen uji yang akan dibandingkan dengan dokumen sumber yang telah tersimpan pada menu laporan. Dokumen uji yang digunakan adalah dokumen yang terindikasi melakukan *plagiarism* dokumen sumber, pengujian yang dilakukan adalah *one to many* yaitu satu dokumen uji dan banyak dokumen sumber. Hasil presentasi antara dokumen uji dan dokumen sumber akan ditampilkan secara berurutan dari persentasi kemiripan tertinggi hingga presentasi kemiripan terendah.



**PENENTUAN TINGKAT KEMIRIPAN LAPORAN KERJA PRAKTEK
MENGUNAKAN METODE PEMBOBOTAN TERM FREQUENCY-INVERSE
DOCUMENT FREQUENCY DAN COSINE SIMILARITY**

Cek Laporan
Dokumen Laporan

☐ 1 kata
 ☐ 2 kata
 ☒ 3 kata
 ☐ 4 kata

☐ Stemming dan Stopword

NO	NIM	NAMA	JUDUL	Tingkat Kemiripan
				<div style="display: flex; justify-content: space-around;"> <div style="text-align: center;"> <div style="width: 70%; height: 10px; background-color: #007bff; margin-bottom: 5px;"></div> <div style="width: 50%; height: 10px; background-color: #007bff; margin-bottom: 5px;"></div> </div> <div style="text-align: center;"> <div style="background-color: #343a40; color: white; padding: 2px 5px;">Rincian</div> <div style="background-color: #343a40; color: white; padding: 2px 5px;">Rincian</div> </div> </div>

Gambar 4.18 Tampilan Cek Laporan