

BAB IV

ANALISA DAN PERANCANGAN

Tahap analisa dan perancangan mempunyai peranan penting pada pembuatan suatu sistem atau aplikasi. Analisa merupakan tahap mengidentifikasi masalah dan kebutuhan yang diperlukan untuk menyelesaikan masalah tersebut. Tahap perancangan yaitu tahap yang dilakukan setelah analisa. Perancangan merupakan tahapan penggambaran sistem yang akan dibuat. Analisa dan Perancangan ini dilakukan untuk menghasilkan sistem yang sesuai kebutuhan

4.1 Analisis Masalah

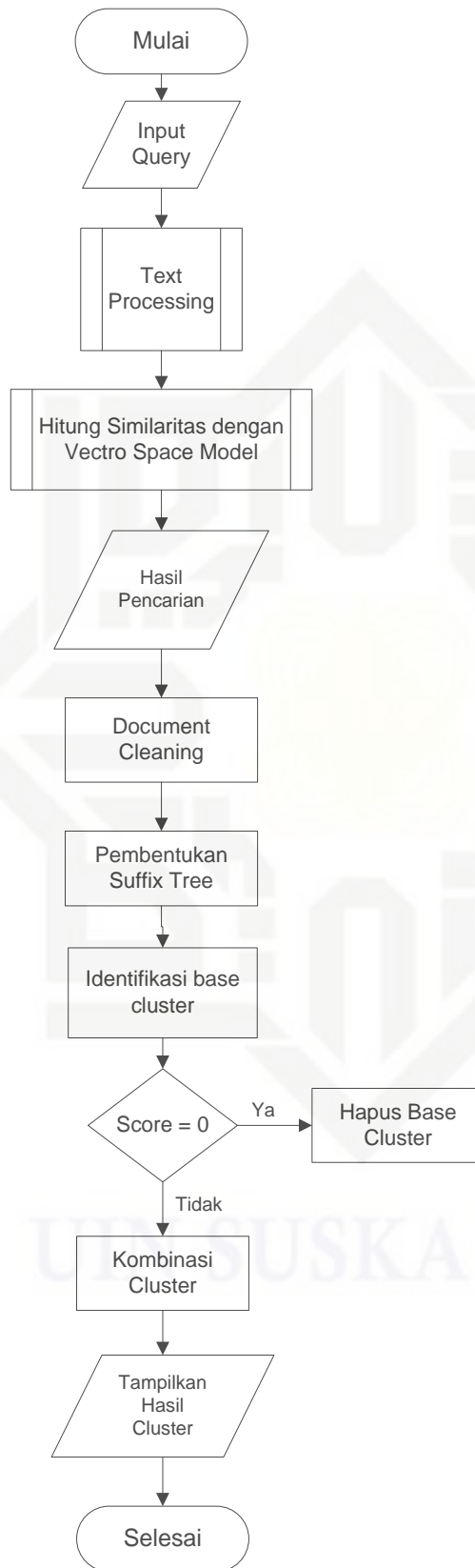
Tugas akhir merupakan salah syarat untuk menyelesaikan studi pada Universitas Sultan Syarif Kasim Riau. Berdasarkan hasil kuisisioner, untuk mendapatkan referensi tugas akhir 98% mahasiswa melakukan pencarian pada *search engine*, 95% mahasiswa tidak mendapatkan hasil pencarian sesuai dengan yang diinginkan dan 71% mahasiswa hanya memperhatikan halaman pertama hasil pencarian.

Berdasarkan permasalahan tersebut maka dirancang sistem temu kembali informasi (*information retrieval*) yang menerapkan pengelompokan pada hasil pencarian (*cluster search result*). Pengelompokan hasil pencarian ini untuk memudahkan pengguna menemukan informasi yang diinginkan pada koleksi dokumen yang besar.

Metode yang digunakan yaitu metode *Vector Space Model* dan *Suffix Tree Clustering*. Metode *vector space model* digunakan untuk mendapatkan hasil pencarian yang relevan dan metode *Suffix Tree Clustering* digunakan untuk pengelompokan hasil pencarian. Berikut gambaran umum sistem:

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 4.1 Gambaran Umum sistem

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 Analisis Sistem

Bagian ini merupakan tahapan analisa terhadap sistem temu kembali informasi dan pengelompokan e-Jurnal. Analisa dilakukan terhadap *input*, *proses* dan *output* sistem.

4.2.1 Analisa Input

Analisa input merupakan analisa terhadap data masukan pada Sistem *Information retrieval*. Data Masukan (*input*) pada sistem yang akan dibangun yaitu:

1. *Corpus*

Corpus merupakan koleksi dokumen yang nantinya akan dijadikan target pencarian. *Corpus* pada penelitian ini yaitu metadata e-Jurnal. Metadata tersebut didapatkan menggunakan API (*Application Programming Interface*) pada website DOAJ (*Directory of Open Access Journal*). *doaj.org* merupakan website penyedia jurnal *open access*. Metadata penelitian ini terdiri dari metadata jurnal *computer science* dan *information technology*. Metadata tersebut terdiri dari:

- a. *Title* : Judul Jurnal
- b. *Author* : Penulis Jurnal
- c. *Abstract* : Abstrak Jurnal
- d. *Affiliation* : Keanggotaan Penulis Jurnal
- e. *Journal Publisher* : Penerbit Jurnal
- f. *Year* : Tahun Jurnal
- g. *ISSN* : Nomor ISSN (*International Standard Serial Number*)
- h. *DOI* : Nomor DOI (*Digital Object Identifier*)
- i. *Keywords* : *Keyword* untuk jurnal
- j. *Link* : *Url* Jurnal tersebut
- k. *Subject* : Subjek Jurnal

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 mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2. Data Administrator

Data ini berisi data-data administrator sistem

3. Data Kumpulan Stop-Word.

Pada sistem ini data *Stop-word* diakses dari website: www.ranks.nl/stopwords

4. Batas Kata

Input jumlah batas kemunculan minimal dan maksimal kata pada keseluruhan dokumen. Jumlah kemunculan kata ini akan mempengaruhi *score base cluster* pada proses *clustering*.

4.2.2 Analisa Proses

Proses merupakan tahapan pengolahan data masukan menjadi *output*.

Proses yang terjadi pada sistem yang akan dibangun yaitu:

1. Proses yang dilakukan oleh admin diantaranya yaitu, proses login, proses *input*, *update* dan *delete corpus*, proses *input*, *update* dan *delete stop-word*, proses *input*, *update* batas kata dan proses *input*, *update* dan *delete* data admin
2. Proses *Text Processing*, ada beberapa tahap yaitu *Tokenizing* (proses pemecahan pemecah kalimat pada dokumen menjadi *term*), *Stop-word Removal* (proses penghapusan kata yang berfrekuensi tinggi), *Stemming* (proses mengubah kata menjadi kata dasar). Pada tahapan ini juga terjadi proses *indexing* dan *term weighting* terhadap *term* menggunakan *TF-IDF*
3. Proses menghitung tingkat relevansi dokumen dengan *query* menggunakan metode *vector space model*
4. Proses *clustering* terhadap hasil pencarian menggunakan metode *suffix tree clustering*. Pada tahapan ini terdapat beberapa proses yaitu, proses *document cleaning* (penghapusan *stop-word* terhadap dokumen), proses identifikasi *base cluster* (pembentukan *suffix tree* dari dokumen serta penghitungan *score base cluster*), proses kombinasi *base cluster* (proses menggabungkan *base cluster* yang mempunyai kesamaan)
5. Penghitungan *Vector Space Model* dan *Suffix Tree Clustering* menggunakan Judul, Abstrak dan *Keyword* E-Jurnal.

6. Proses pencarian jurnal yang dilakukan oleh *user* dengan memasukkan *query*

4.2.3 Analisa Output

Output yang diharapkan dari sistem temu kembali informasi dan pengelompokan e-Jurnal ini yaitu *user* mendapatkan daftar hasil pencarian beserta kelompok hasil pencarian tersebut.

4.3 Analisis Proses Information Retrieval

Dalam membangun sistem temu kembali informasi (*information retrieval*) ada beberapa proses yaitu *text-preprocessing* dan menghitung tingkat relevan dokumen dengan *query* serta perangkan dengan *Vector Space Model*.

4.3.1 Text Processing

Tahap ini melakukan pemrosesan terhadap teks. Proses ini bertujuan untuk mempersiapkan teks menjadi data yang akan mengalami proses lebih lanjut. Tujuan pemrosesan teks (*text processing*) adalah mengubah banyak bentuk kata-kata yang dapat terjadi menjadi *index term* yang lebih konsisten. *Index term* adalah representasi dari isi dokumen yang digunakan untuk pencarian (Croft et al, 2015). Pada tahap ini pada umumnya terdapat beberapa proses, antara lain *tokenizing*, penghapusan *stop-word stemming*, dan *term weighting*. Berikut contoh pencarian hasil relevan dengan *query* “*Recognition Using Backpropagation*”.

Contoh Kasus:

Tabel 4.1 Contoh Kasus

No Dokumen	Judul
D1	A Robust Deep-Learning-Based Detector for Real-Time Tomato Plant Diseases and Pests Recognition
D2	The Prediction of Bandwidth On Need Computer Network Through Artificial Neural Network Method of Backpropagation
D3	Multi-Digit Handwritten Sindhi Numerals Recognition using SOM Neural Network
D4	Alphabet Sign Language Recognition Using Leap Motion Technology And Rule Based Backpropagation-Genetic Algorithm Neural Network (RBBPGANN)
D5	An Adaboost-Backpropagation Neural Network for Automated Image Sentiment Classification
D6	Spoof Detection for Finger-Vein Recognition System Using NIR Camera
D7	Prediction of Negative Conversion Days of Childhood Nephrotic Syndrome Based on the Improved Backpropagation Neural Network with Momentum

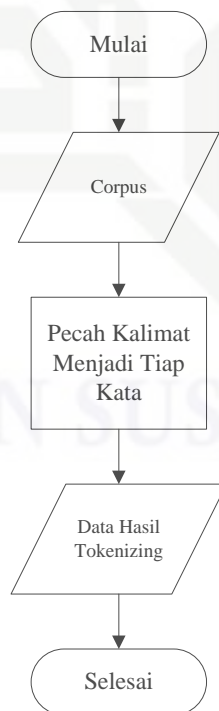
Hak Cipta Dilindungi Undang-Undang

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 mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

D8	Knowledge Mining from Clinical Datasets Using Rough Sets and Backpropagation Neural Network
D9	Impact of Ethnic Group on Human Emotion Recognition Using Backpropagation Neural Network
D10	An Adaptive Filtering Algorithm Based on Genetic Algorithm-Backpropagation Network
D11	The application of backpropagation neural network method to estimate the sediment loads
D12	Static Hand Gesture Recognition of Indonesian Sign Language System Based on Backpropagation Neural Networks
D13	Facial Expression Recognition By Using Fisherface Methode With Backpropagation Neural Network
D14	facial Expression Recognition By Using Fisherface Methode With Backpropagation Neural Network
D15	Obstacle Recognition Based on Machine Learning for On-Chip LiDAR Sensors in a Cyber-Physical System
Query	Recognition Using Backpropagation

1. Tokenizing

Tokenizing merupakan proses pemecahan kalimat menjadi *term*. Proses ini juga menghilangkan tanda baca dan merubah semua karakter menjadi huruf kecil (*Lower case*). Tujuan *tokenizing* adalah eksplorasi kata-kata dalam sebuah kalimat. Daftar token menjadi masukan untuk diproses lebih lanjut.



Gambar 4.2 Flowchart Tokenizing

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.

Tabel 4.2 Hasil Tokenizing

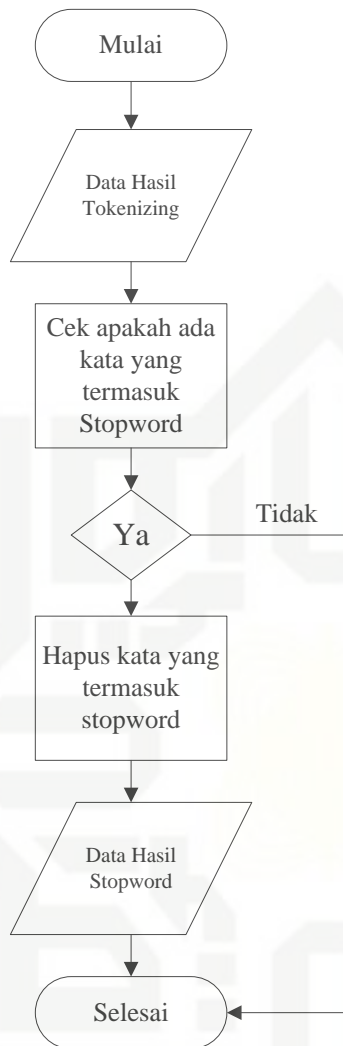
No Dokumen	Judul
D1	a robust deep learning based detector for real time tomato plant diseases and pests recognition
D2	the prediction of bandwidth on need computer network through artificial neural network method of backpropagation
D3	multi digit handwritten sindhi numerals recognition using som neural network
D4	alphabet sign language recognition using leap motion technology and rule based backpropagation genetic algorithm neural network rbbpgann
D5	an adaboost backpropagation neural network for automated image sentiment classification
D6	spoof detection for finger vein recognition system using nir camera
D7	prediction of negative conversion days of childhood nephrotic syndrome based on the improved backpropagation neural network with momentum
D8	knowledge mining from clinical datasets using rough sets and backpropagation neural network
D9	impact of ethnic group on human emotion recognition using backpropagation neural network
D10	an adaptive filtering algorithm based on genetic algorithm backpropagation network
D11	the application of backpropagation neural network method to estimate the sediment loads
D12	static hand gesture recognition of indonesian sign language system based on backpropagation neural networks
D13	facial expression recognition by using fisherface methode with backpropagation neural network
D14	online handwritten character recognition using an optical backpropagation neural network
D15	obstacle recognition based on machine learning for on chip lidar sensors in a cyber physical system
Query	recognition using backpropagation

2. Penghapusan Stop-Word

Pada dokumen terdapat banyak kata yang berulang yang pada dasarnya tidak memiliki makna. Pada sistem *information retrieval* kata-kata tersebut disebut *stop-word*. Disebut *stop-word* karena *text processing* akan berhenti saar ada kata tersebut dan dihapus. Penghapusan *stop-word* dapat menurunkan ukuran indeks, meningkatkan efisiensi dan keefektifan pengambilan informasi (Croft et al, 2015). Pada penelitian ini data *stop-word* yang digunakan dari website: www.ranks.nl/stopwords yang terdapat pada Lampiran IB

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 4.3 Flowchart Stop-word Removal

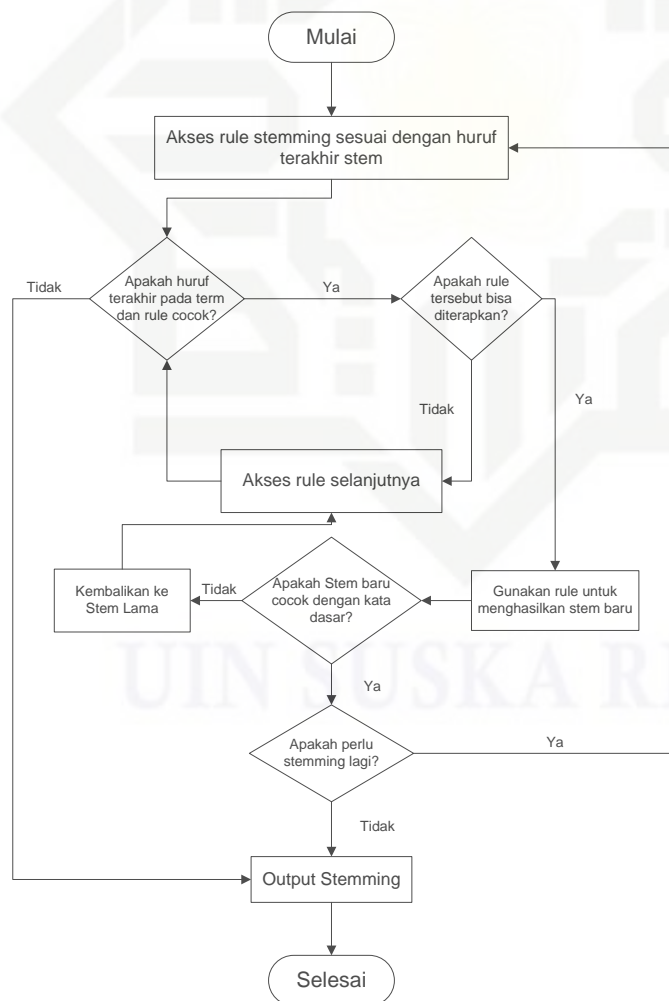
Tabel 4.3 Hasil penghapusan Stop-word

No Dokumen	Judul
D1	robust deep learning based detector real time tomato plant diseases pests recognition
D2	prediction bandwidth need computer network artificial neural network method backpropagation
D3	multi digit handwritten sindhi numerals recognition som neural network
D4	alphabet sign language recognition leap motion technology rule based backpropagation genetic algorithm neural network rbbpgann
D5	adaboost backpropagation neural network automated image sentiment classification
D6	spooof detection finger vein recognition system nir camera
D7	prediction negative conversion days of childhood nephrotic syndrome based improved backpropagation neural network momentum
D8	knowledge mining clinical datasets rough sets backpropagation neural network

D9	impact ethnic group human emotion recognition backpropagation neural network
D10	adaptive filtering algorithm based genetic algorithm backpropagation network
D11	application backpropagation neural network method estimate sediment loads
D12	static hand gesture recognition indonesian sign language system based backpropagation neural networks
D13	facial expression recognition fisherface method backpropagation neural network
D14	online handwritten character recognition optical backpropagation neural network
D15	obstacle recognition based machine learning chip lidar sensors in a cyber physical system
Query	recognition backpropagation

3. Stemming

Stemming merupakan proses mengubah kata menjadi kata dasar. Pada tahap ini *stemming* menggunakan *Porter Stemmer* untuk menghilangkan *suffix* (akhiran) pada *term* Bahasa Inggris. Berikut *flowchart porter stemmer*.



Gambar 4.4 Flowchart *Stemming*

- 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.

Tabel 4.4 Hasil Stemming

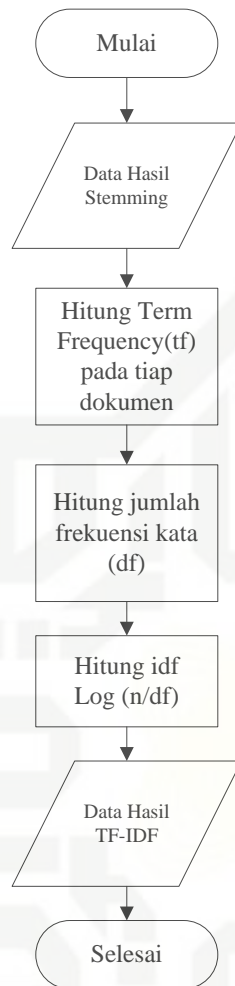
No Dokumen	Judul
D1	robust deep learn base detector real time tomato plant disease pest recognit
D2	predict bandwidth computer network artificial neural network method backpropagation
D3	multi digit handwritten sindhi numeral recognit som neural network
D4	alphabet sign language recognit leap motion teknologi rule base backpropagation genetic algorithm neural network rbbpgan
D5	adaboost backpropagation neural network automated image sentiment classific
D6	spooof detect finger vein recognit system nir camera
D7	predict negative convers day childhood nephrotic syndrome base Improv backpropagation neural network momentum
D8	knowledge mining clinic dataset rough set backpropagation neural network
D9	impact ethnic group human emotion recognit backpropagation neural network
D10	adapt filter algorithm base genetic algorithm backpropagation network
D11	application backpropagation neural network method estim sediment load
D12	static hand gesture recognit indonesian sign language system base backpropagation neural network
D13	facial expression recognit fisherface method backpropagation neural network
D14	online handwritten character recognit optical backpropagation neural network
D15	obstacle recognit base machine learn chip lidar sensor cyber physic system
Query	recognit backpropagation

4. Term Weighting

Pada tahap ini dilakukan proses pembuatan *indexing* terhadap *term* yang ada pada setiap dokumen. Selanjutnya dilakukan *term weighting* atau pembobotan terhadap setiap *term* sesuai dengan frekuensi kemunculan. Pembobotan dilakukan dengan *TF-IDF* (*term frequency-inverse document frequency*). Berikut *flowchart 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.



Gambar 4.5 Flowchart TF-IDF

Berikut adalah keterangan tabel perhitungan TF-IDF:

- a. Tabel 4.5 yaitu tabel perhitungan TF (*Term Frequency*)
- b. Tabel 4.6 yaitu tabel perhitungan TF-IDF (*Term Frequency-Inverse Document Frequency*)

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, pen-
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin

Tabel 4.5 Perhitungan TF (Term Frequency)

Term	Query	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	df	D/df	IDF
recognit	1	1	0	1	1	0	1	0	0	1	0	0	1	1	1	1	10	1.5	0.176
backpropagation	1	0	1	0	1	1	0	1	1	1	1	1	1	1	1	0	12	1.25	0.097
robust	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	15	1.176
deep	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	15	1.176
learn	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	7.5	0.875
base	0	1	0	0	1	0	0	1	0	0	1	0	1	0	0	1	6	2.5	0.398
detector	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	15	1.176
real	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	15	1.176
time	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	15	1.176
tomato	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	15	1.176
plant	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	15	1.176
disease	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	15	1.176
pest	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	15	1.176
predict	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	2	7.5	0.875
bandwidth	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	15	1.176
computer	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	15	1.176
network	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1	0	12	1.25	0.097
artificial	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	15	1.176
neural	0	0	1	1	1	1	0	1	1	1	0	1	1	1	1	0	11	1.364	0.135
method	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	3	5	0.699
multi	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	15	1.176
digit	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	15	1.176
handwritten	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	2	7.5	0.875



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, pen-
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin

Term	Query	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	df	D/df	IDF
sindhi	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	15	1.176
numeral	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	15	1.176
som	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	15	1.176
alphabet	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	15	1.176
sign	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	2	7.5	0.875
language	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	2	7.5	0.875
leap	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	15	1.176
motion	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	15	1.176
teknologi	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	15	1.176
rule	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	15	1.176
genetic	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	2	7.5	0.875
algorithm	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	2	7.5	0.875
rbbpgan	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	15	1.176
adaboost	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	15	1.176
automated	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	15	1.176
image	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	15	1.176
sentiment	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	15	1.176
classific	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	15	1.176
spooof	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	15	1.176
detect	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	15	1.176
finger	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	15	1.176
vein	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	15	1.176
system	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	3	5	0.699
nir	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	15	1.176



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, pen-
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin

Term	Query	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	df	D/df	IDF
camera	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	15	1.176
negative	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	15	1.176
convers	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	15	1.176
day	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	15	1.176
childhood	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	15	1.176
nephrotic	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	15	1.176
syndrome	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	15	1.176
improv	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	15	1.176
momentum	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	15	1.176
knowledge	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	15	1.176
mining	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	15	1.176
clinic	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	15	1.176
dataset	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	15	1.176
rough	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	15	1.176
set	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	15	1.176
impact	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	15	1.176
ethnic	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	15	1.176
group	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	15	1.176
human	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	15	1.176
emotion	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	15	1.176
adapt	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	15	1.176
filter	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	15	1.176
application	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	15	1.176
estim	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	15	1.176



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, pen-
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin

Term	Query	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	df	D/df	IDF
sediment	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	15	1.176
load	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	15	1.176
static	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	15	1.176
hand	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	15	1.176
gesture	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	15	1.176
indonesian	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	15	1.176
facial	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	15	1.176
expression	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	15	1.176
fisherface	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	15	1.176
online	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	15	1.176
character	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	15	1.176
optical	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	15	1.176
obstacle	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	15	1.176
machine	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	15	1.176
chip	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	15	1.176
lidar	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	15	1.176
sensor	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	15	1.176
cyber	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	15	1.176
physic	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	15	1.176

Tabel 4.6 Perhitungan *TF-IDF*

Term	Query	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15
recognit	0.176	0.176	0	0.176	0.176	0	0.176	0	0	0.176	0	0	0.176	0.176	0.176	0.176
backpropagation	0.097	0	0.097	0	0.097	0.097	0	0.097	0.097	0.097	0.097	0.097	0.097	0.097	0.097	0
robust	0	1.176	0	0	0	0	0	0	0	0	0	0	0	0	0	0
deep	0	1.176	0	0	0	0	0	0	0	0	0	0	0	0	0	0
learn	0	0.875	0	0	0	0	0	0	0	0	0	0	0	0	0	0.875
base	0	0.398	0	0	0.398	0	0	0.398	0	0	0.398	0	0.398	0	0	0.398
detector	0	1.176	0	0	0	0	0	0	0	0	0	0	0	0	0	0
real	0	1.176	0	0	0	0	0	0	0	0	0	0	0	0	0	0
time	0	1.176	0	0	0	0	0	0	0	0	0	0	0	0	0	0
tomato	0	1.176	0	0	0	0	0	0	0	0	0	0	0	0	0	0
plant	0	1.176	0	0	0	0	0	0	0	0	0	0	0	0	0	0
disease	0	1.176	0	0	0	0	0	0	0	0	0	0	0	0	0	0
pest	0	1.176	0	0	0	0	0	0	0	0	0	0	0	0	0	0
predict	0	0	0.875	0	0	0	0	0.875	0	0	0	0	0	0	0	0
bandwidth	0	0	1.176	0	0	0	0	0	0	0	0	0	0	0	0	0
computer	0	0	1.176	0	0	0	0	0	0	0	0	0	0	0	0	0
network	0	0	0.097	0.097	0.097	0.097	0	0.097	0.097	0.097	0.097	0.097	0.097	0.097	0.097	0
artificial	0	0	1.176	0	0	0	0	0	0	0	0	0	0	0	0	0
neural	0	0	0.135	0.135	0.135	0.135	0	0.135	0.135	0.135	0	0.135	0.135	0.135	0.135	0
method	0	0	0.699	0	0	0	0	0	0	0	0	0.699	0	0.699	0	0
multi	0	0	0	1.176	0	0	0	0	0	0	0	0	0	0	0	0
digit	0	0	0	1.176	0	0	0	0	0	0	0	0	0	0	0	0
handwritten	0	0	0	0.875	0	0	0	0	0	0	0	0	0	0	0.875	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, pen-
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izn

Term	Query	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15
sindhi	0	0	0	1.176	0	0	0	0	0	0	0	0	0	0	0	0
numeral	0	0	0	1.176	0	0	0	0	0	0	0	0	0	0	0	0
som	0	0	0	1.176	0	0	0	0	0	0	0	0	0	0	0	0
alphabet	0	0	0	0	1.176	0	0	0	0	0	0	0	0	0	0	0
sign	0	0	0	0	0.875	0	0	0	0	0	0	0	0.875	0	0	0
language	0	0	0	0	0.875	0	0	0	0	0	0	0	0.875	0	0	0
leap	0	0	0	0	1.176	0	0	0	0	0	0	0	0	0	0	0
motion	0	0	0	0	1.176	0	0	0	0	0	0	0	0	0	0	0
tehnologi	0	0	0	0	1.176	0	0	0	0	0	0	0	0	0	0	0
rule	0	0	0	0	1.176	0	0	0	0	0	0	0	0	0	0	0
genetic	0	0	0	0	0.875	0	0	0	0	0	0.875	0	0	0	0	0
algorithm	0	0	0	0	0.875	0	0	0	0	0	0.875	0	0	0	0	0
rbbpgan	0	0	0	0	1.176	0	0	0	0	0	0	0	0	0	0	0
adaboost	0	0	0	0	0	1.176	0	0	0	0	0	0	0	0	0	0
automated	0	0	0	0	0	1.176	0	0	0	0	0	0	0	0	0	0
image	0	0	0	0	0	1.176	0	0	0	0	0	0	0	0	0	0
sentiment	0	0	0	0	0	1.176	0	0	0	0	0	0	0	0	0	0
classific	0	0	0	0	0	1.176	0	0	0	0	0	0	0	0	0	0
spooof	0	0	0	0	0	0	1.176	0	0	0	0	0	0	0	0	0
detect	0	0	0	0	0	0	1.176	0	0	0	0	0	0	0	0	0
finger	0	0	0	0	0	0	1.176	0	0	0	0	0	0	0	0	0
vein	0	0	0	0	0	0	1.176	0	0	0	0	0	0	0	0	0
system	0	0	0	0	0	0	0.699	0	0	0	0	0	0.699	0	0	0.699
nir	0	0	0	0	0	0	1.176	0	0	0	0	0	0	0	0	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, pen-
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin

Term	Query	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15
camera	0	0	0	0	0	0	1.176	0	0	0	0	0	0	0	0	0
negative	0	0	0	0	0	0	0	1.176	0	0	0	0	0	0	0	0
convers	0	0	0	0	0	0	0	1.176	0	0	0	0	0	0	0	0
day	0	0	0	0	0	0	0	1.176	0	0	0	0	0	0	0	0
childhood	0	0	0	0	0	0	0	1.176	0	0	0	0	0	0	0	0
nephrotic	0	0	0	0	0	0	0	1.176	0	0	0	0	0	0	0	0
syndrome	0	0	0	0	0	0	0	1.176	0	0	0	0	0	0	0	0
improv	0	0	0	0	0	0	0	1.176	0	0	0	0	0	0	0	0
momentum	0	0	0	0	0	0	0	1.176	0	0	0	0	0	0	0	0
knowledge	0	0	0	0	0	0	0	0	1.176	0	0	0	0	0	0	0
mining	0	0	0	0	0	0	0	0	1.176	0	0	0	0	0	0	0
clinic	0	0	0	0	0	0	0	0	1.176	0	0	0	0	0	0	0
dataset	0	0	0	0	0	0	0	0	1.176	0	0	0	0	0	0	0
rough	0	0	0	0	0	0	0	0	1.176	0	0	0	0	0	0	0
set	0	0	0	0	0	0	0	0	1.176	0	0	0	0	0	0	0
impact	0	0	0	0	0	0	0	0	0	1.176	0	0	0	0	0	0
ethnic	0	0	0	0	0	0	0	0	0	1.176	0	0	0	0	0	0
group	0	0	0	0	0	0	0	0	0	1.176	0	0	0	0	0	0
human	0	0	0	0	0	0	0	0	0	1.176	0	0	0	0	0	0
emotion	0	0	0	0	0	0	0	0	0	1.176	0	0	0	0	0	0
adapt	0	0	0	0	0	0	0	0	0	0	1.176	0	0	0	0	0
filter	0	0	0	0	0	0	0	0	0	0	1.176	0	0	0	0	0
application	0	0	0	0	0	0	0	0	0	0	0	1.176	0	0	0	0
estim	0	0	0	0	0	0	0	0	0	0	0	1.176	0	0	0	0



Term	Query	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15
sediment	0	0	0	0	0	0	0	0	0	0	0	1.176	0	0	0	0
load	0	0	0	0	0	0	0	0	0	0	0	1.176	0	0	0	0
static	0	0	0	0	0	0	0	0	0	0	0	0	1.176	0	0	0
hand	0	0	0	0	0	0	0	0	0	0	0	0	1.176	0	0	0
gesture	0	0	0	0	0	0	0	0	0	0	0	0	1.176	0	0	0
indonesian	0	0	0	0	0	0	0	0	0	0	0	0	1.176	0	0	0
facial	0	0	0	0	0	0	0	0	0	0	0	0	0	1.176	0	0
expression	0	0	0	0	0	0	0	0	0	0	0	0	0	1.176	0	0
fisherface	0	0	0	0	0	0	0	0	0	0	0	0	0	1.176	0	0
online	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.176	0
character	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.176	0
optical	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.176	0
obstacle	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.176
machine	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.176
chip	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.176
lidar	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.176
sensor	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.176
cyber	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.176
physic	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.176

Hak Cipta Dilindungi Undang-Undang

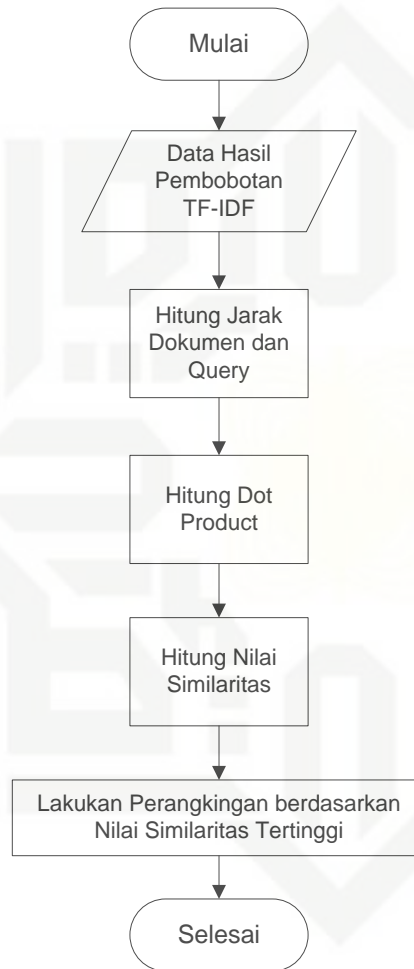
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, pen-
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izi-

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.

4.3.2 Vector Space Model

Setelah proses *term weighting* (pembobotan *term*) selesai dilakukan, maka tahapan selanjutnya yaitu menghitung tingkat relevansi antara *query* dengan dokumen menggunakan *vector space mode* untuk mendapatkan hasil pencarian relevan. Berikut *flowchart vector space model*:



Gambar 4.6 *Flowchart Vector Space Model*

Berikut adalah keterangan table perhitungan *Vector Space Model*:

- a. Tabel 4.7 perhitungan jarak *query* dan *document*
- b. Tabel 4.8 perhitungan dot produk



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, pen-
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin

Tabel 4.7 Perhitungan Jarak query dan dokumen.

Term	Query	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15
recognit	0.031	0.031	0	0.031	0.031	0	0.031	0	0	0.031	0	0	0.031	0.031	0.031	0.031
backpropagation	0.009	0	0.009	0	0.009	0.009	0	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0
robust	0	1.383	0	0	0	0	0	0	0	0	0	0	0	0	0	0
deep	0	1.383	0	0	0	0	0	0	0	0	0	0	0	0	0	0
learn	0	0.766	0	0	0	0	0	0	0	0	0	0	0	0	0	0.766
base	0	0.158	0	0	0.158	0	0	0.158	0	0	0.158	0	0.158	0	0	0.158
detector	0	1.383	0	0	0	0	0	0	0	0	0	0	0	0	0	0
real	0	1.383	0	0	0	0	0	0	0	0	0	0	0	0	0	0
time	0	1.383	0	0	0	0	0	0	0	0	0	0	0	0	0	0
tomato	0	1.383	0	0	0	0	0	0	0	0	0	0	0	0	0	0
plant	0	1.383	0	0	0	0	0	0	0	0	0	0	0	0	0	0
disease	0	1.383	0	0	0	0	0	0	0	0	0	0	0	0	0	0
pest	0	1.383	0	0	0	0	0	0	0	0	0	0	0	0	0	0
predict	0	0	0.766	0	0	0	0	0.766	0	0	0	0	0	0	0	0
bandwidth	0	0	1.383	0	0	0	0	0	0	0	0	0	0	0	0	0
computer	0	0	1.383	0	0	0	0	0	0	0	0	0	0	0	0	0
network	0	0	0.009	0.009	0.009	0.009	0	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0
artificial	0	0	1.383	0	0	0	0	0	0	0	0	0	0	0	0	0
neural	0	0	0.018	0.018	0.018	0.018	0	0.018	0.018	0.018	0	0.018	0.018	0.018	0.018	0
method	0	0	0.489	0	0	0	0	0	0	0	0	0.489	0	0.489	0	0
multi	0	0	0	1.383	0	0	0	0	0	0	0	0	0	0	0	0
digit	0	0	0	1.383	0	0	0	0	0	0	0	0	0	0	0	0
handwritten	0	0	0	0.766	0	0	0	0	0	0	0	0	0	0	0.766	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, pen-
gumpulan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 - b. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izn.

Term	Query	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15
sindhi	0	0	0	1.383	0	0	0	0	0	0	0	0	0	0	0	0
numeral	0	0	0	1.383	0	0	0	0	0	0	0	0	0	0	0	0
som	0	0	0	1.383	0	0	0	0	0	0	0	0	0	0	0	0
alphabet	0	0	0	0	1.383	0	0	0	0	0	0	0	0	0	0	0
sign	0	0	0	0	0.766	0	0	0	0	0	0	0	0.766	0	0	0
language	0	0	0	0	0.766	0	0	0	0	0	0	0	0.766	0	0	0
leap	0	0	0	0	1.383	0	0	0	0	0	0	0	0	0	0	0
motion	0	0	0	0	1.383	0	0	0	0	0	0	0	0	0	0	0
tehnologi	0	0	0	0	1.383	0	0	0	0	0	0	0	0	0	0	0
rule	0	0	0	0	1.383	0	0	0	0	0	0	0	0	0	0	0
genetic	0	0	0	0	0.766	0	0	0	0	0	0.766	0	0	0	0	0
algorithm	0	0	0	0	0.766	0	0	0	0	0	0.766	0	0	0	0	0
rbbpgan	0	0	0	0	1.383	0	0	0	0	0	0	0	0	0	0	0
adaboost	0	0	0	0	0	1.383	0	0	0	0	0	0	0	0	0	0
automated	0	0	0	0	0	1.383	0	0	0	0	0	0	0	0	0	0
image	0	0	0	0	0	1.383	0	0	0	0	0	0	0	0	0	0
sentiment	0	0	0	0	0	1.383	0	0	0	0	0	0	0	0	0	0
classific	0	0	0	0	0	1.383	0	0	0	0	0	0	0	0	0	0
spooof	0	0	0	0	0	0	1.383	0	0	0	0	0	0	0	0	0
detect	0	0	0	0	0	0	1.383	0	0	0	0	0	0	0	0	0
finger	0	0	0	0	0	0	1.383	0	0	0	0	0	0	0	0	0
vein	0	0	0	0	0	0	1.383	0	0	0	0	0	0	0	0	0
system	0	0	0	0	0	0	0.489	0	0	0	0	0	0.489	0	0	0.489
nir	0	0	0	0	0	0	1.383	0	0	0	0	0	0	0	0	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, pen-
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin

Term	Query	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15
camera	0	0	0	0	0	0	1.383	0	0	0	0	0	0	0	0	0
negative	0	0	0	0	0	0	0	1.383	0	0	0	0	0	0	0	0
convers	0	0	0	0	0	0	0	1.383	0	0	0	0	0	0	0	0
day	0	0	0	0	0	0	0	1.383	0	0	0	0	0	0	0	0
childhood	0	0	0	0	0	0	0	1.383	0	0	0	0	0	0	0	0
nephrotic	0	0	0	0	0	0	0	1.383	0	0	0	0	0	0	0	0
syndrome	0	0	0	0	0	0	0	1.383	0	0	0	0	0	0	0	0
improv	0	0	0	0	0	0	0	1.383	0	0	0	0	0	0	0	0
momentum	0	0	0	0	0	0	0	1.383	0	0	0	0	0	0	0	0
knowledge	0	0	0	0	0	0	0	0	1.383	0	0	0	0	0	0	0
mining	0	0	0	0	0	0	0	0	1.383	0	0	0	0	0	0	0
clinic	0	0	0	0	0	0	0	0	1.383	0	0	0	0	0	0	0
dataset	0	0	0	0	0	0	0	0	1.383	0	0	0	0	0	0	0
rough	0	0	0	0	0	0	0	0	1.383	0	0	0	0	0	0	0
set	0	0	0	0	0	0	0	0	1.383	0	0	0	0	0	0	0
impact	0	0	0	0	0	0	0	0	0	1.383	0	0	0	0	0	0
ethnic	0	0	0	0	0	0	0	0	0	1.383	0	0	0	0	0	0
group	0	0	0	0	0	0	0	0	0	1.383	0	0	0	0	0	0
human	0	0	0	0	0	0	0	0	0	1.383	0	0	0	0	0	0
emotion	0	0	0	0	0	0	0	0	0	1.383	0	0	0	0	0	0
adapt	0	0	0	0	0	0	0	0	0	0	1.383	0	0	0	0	0
filter	0	0	0	0	0	0	0	0	0	0	1.383	0	0	0	0	0
application	0	0	0	0	0	0	0	0	0	0	0	1.383	0	0	0	0
estim	0	0	0	0	0	0	0	0	0	0	0	1.383	0	0	0	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, pen-
gumpulan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 - b. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izi-

Term	Query	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15
sediment	0	0	0	0	0	0	0	0	0	0	0	1.383	0	0	0	0
load	0	0	0	0	0	0	0	0	0	0	0	1.383	0	0	0	0
static	0	0	0	0	0	0	0	0	0	0	0	0	1.383	0	0	0
hand	0	0	0	0	0	0	0	0	0	0	0	0	1.383	0	0	0
gesture	0	0	0	0	0	0	0	0	0	0	0	0	1.383	0	0	0
indonesian	0	0	0	0	0	0	0	0	0	0	0	0	1.383	0	0	0
facial	0	0	0	0	0	0	0	0	0	0	0	0	0	1.383	0	0
expression	0	0	0	0	0	0	0	0	0	0	0	0	0	1.383	0	0
fisherface	0	0	0	0	0	0	0	0	0	0	0	0	0	1.383	0	0
online	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.383	0
character	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.383	0
optical	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.383	0
obstacle	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.383
machine	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.383
chip	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.383
lidar	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.383
sensor	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.383
cyber	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.383
physic	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.383
Jumlah	0.04	13.402	5.44	7.739	11.587	6.951	8.818	12.024	8.334	6.982	4.474	6.057	7.778	4.705	4.982	11.125
Sqrt (Jumlah)	0.2	3.661	2.332	2.782	3.404	2.636	2.97	3.468	2.887	2.642	2.115	2.461	2.789	2.169	2.232	3.335

Tabel 4.8 Perhitungan Dot Produk

Term	Q*D1	Q*D2	Q*D3	Q*D4	Q*D5	Q*D6	Q*D7	Q*D8	Q*D9	Q*D10	Q*D11	Q*D12	Q*D13	Q*D14	Q*D15
recognit	0.000961	0	0.000961	0.000961	0	0.000961	0	0	0.000961	0	0	0.000961	0.000961	0.000961	0.000961
backpropagation	0	0.000081	0	0.000081	0.000081	0	0.000081	0.000081	0.000081	0.000081	0.000081	0.000081	0.000081	0.000081	0
robust	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
deep	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
learn	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
base	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
detector	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
real	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
time	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
tomato	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
plant	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
disease	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
pest	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
predict	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
bandwidth	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
computer	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
network	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
artificial	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
neural	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
method	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
multi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
digit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
handwritten	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0



Hak Cipta Dilindungi Undang-Undang
 1. Dilarang mengutip, memperbanyak, atau menyebarkan dalam bentuk apapun tanpa izin dari penerbit.
 2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari penerbit.

© Hak Cipta milik UIN Suska Riau
 State Islamic University

Term	Q*D1	Q*D2	Q*D3	Q*D4	Q*D5	Q*D6	Q*D7	Q*D8	Q*D9	Q*D10	Q*D11	Q*D12	Q*D13	Q*D14	Q*D15
sindhi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
numeral	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
som	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
alphabet	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
sign	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
language	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
leap	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
motion	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
teknologi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
rule	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
genetic	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
algorithm	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
rbbpgan	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
adaboost	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
automated	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
image	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
sentiment	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
classific	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
spoof	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
detect	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
finger	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
vein	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
system	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
nir	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0



Hak Cipta Dilindungi Undang-Undang
 1. Dilarang mengutip atau menyalin sebagian atau seluruh karya tulis ini tanpa izin dari penerbit.
 2. Dilarang mengutip atau menyalin sebagian atau seluruh karya tulis ini tanpa izin dari penerbit.

© Hak Cipta milik UIN Suska Riau

Term	Q*D1	Q*D2	Q*D3	Q*D4	Q*D5	Q*D6	Q*D7	Q*D8	Q*D9	Q*D10	Q*D11	Q*D12	Q*D13	Q*D14	Q*D15
camera	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
negative	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
convers	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
day	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
childhood	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
nephrotic	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
syndrome	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
improv	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
momentum	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
knowledge	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
mining	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
clinic	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
dataset	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
rough	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
set	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
impact	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ethnic	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
group	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
human	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
emotion	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
adapt	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
filter	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
application	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
estim	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0



Hak Cipta Dilindungi Undang-Undang
 1. Dilarang mengutip, sebagian atau seluruhnya, tanpa menyebutkan sumber:
 a. Untuk keperluan lain yang bersifat komersial, termasuk iklan atau promosi lainnya, penyusunan laporan, pen-
 b. Pengutipan tidak diperbolehkan untuk tujuan komersial atau untuk tujuan lain yang melanggar hak cipta UIN Suska Riau.
 2. Dilarang mengutip atau menyalin sebagian atau seluruhnya dalam bentuk apapun tanpa izin

© Hak cipta milik UIN Suska Riau

State Islamic Univerr



Term	Q*D1	Q*D2	Q*D3	Q*D4	Q*D5	Q*D6	Q*D7	Q*D8	Q*D9	Q*D10	Q*D11	Q*D12	Q*D13	Q*D14	Q*D15
sediment	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
load	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
static	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
hand	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
gesture	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
indonesian	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
facial	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
expression	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
fisherface	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
online	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
character	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
optical	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
obstacle	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
machine	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
chip	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
lidar	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
sensor	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
cyber	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
physic	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Jumlah	0.000961	0.000081	0.000961	0.001042	0.000081	0.000961	0.000081	0.000081	0.001042	0.000081	0.000081	0.001042	0.001042	0.001042	0.000961

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.

Setelah mendapatkan nilai jarak antara *query* dan dokumen, dan nilai dot produk, maka proses selanjutnya yang dilakukan yaitu menghitung *cosine similarity* (menghitung nilai kosinus sudut antara dua vektor). Berikut perhitungan *cosine similarity*:

$$D1 = \frac{q \cdot d_j}{|q| \cdot |d_j|} = \frac{0.000961}{0.2 \cdot 3.661} = 0.001312483$$

$$D2 = \frac{q \cdot d_j}{|q| \cdot |d_j|} = \frac{0.000081}{0.2 \cdot 2.332} = 0.000173671$$

$$D3 = \frac{q \cdot d_j}{|q| \cdot |d_j|} = \frac{0.000961}{0.2 \cdot 2.782} = 0.001727175$$

$$D3 = \frac{q \cdot d_j}{|q| \cdot |d_j|} = \frac{0.000961}{0.2 \cdot 2.782} = 0.001727175$$

$$D4 = \frac{q \cdot d_j}{|q| \cdot |d_j|} = \frac{0.001042}{0.2 \cdot 3.404} = 0.001530552$$

$$D5 = \frac{q \cdot d_j}{|q| \cdot |d_j|} = \frac{0.000081}{0.2 \cdot 2.636} = 0.000153642$$

$$D6 = \frac{q \cdot d_j}{|q| \cdot |d_j|} = \frac{0.000961}{0.2 \cdot 2.97} = 0.001617845$$

$$D7 = \frac{q \cdot d_j}{|q| \cdot |d_j|} = \frac{0.000081}{0.2 \cdot 3.468} = 0.000116782$$

$$D8 = \frac{q \cdot d_j}{|q| \cdot |d_j|} = \frac{0.000081}{0.2 \cdot 2.887} = 0.000140284$$

$$D9 = \frac{q \cdot d_j}{|q| \cdot |d_j|} = \frac{0.001042}{0.2 \cdot 2.642} = 0.001971991$$

$$D10 = \frac{q \cdot d_j}{|q| \cdot |d_j|} = \frac{0.000081}{0.2 \cdot 2.115} = 0.000191489$$

$$D11 = \frac{q \cdot d_j}{|q| \cdot |d_j|} = \frac{0.000081}{0.2 \cdot 2.461} = 0.000164567$$

$$D12 = \frac{q \cdot d_j}{|q| \cdot |d_j|} = \frac{0.001042}{0.2 \cdot 2.789} = 0.001868053$$

$$D13 = \frac{q \cdot d_j}{|q| \cdot |d_j|} = \frac{0.001042}{0.2 \cdot 2.169} = 0.002402029$$

$$D14 = \frac{q \cdot d_j}{|q| \cdot |d_j|} = \frac{0.001042}{0.2 \cdot 2.232} = 0.002334229$$

$$D15 = \frac{q \cdot d_j}{|q| \cdot |d_j|} = \frac{0.000961}{0.2 \cdot 3.335} = 0.00144078$$

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.

Berdasarkan hasil perhitungan *vector space model* maka diperoleh hasil untuk 15 dokumen tersebut:

Tabel 4.9 Hasil Pencarian Relevan

Ranking	NO Dokumen	Nilai Similaritas
1	D13	0.002402029
2	D14	0.002334229
3	D9	0.001971991
4	D12	0.001868053
5	D3	0.001727175
6	D6	0.001617845
7	D4	0.001530552
8	D15	0.00144078
9	D1	0.001312483
10	D10	0.000191489
11	D2	0.000173671
12	D11	0.000164567
13	D5	0.000153642
14	D8	0.000140284
15	D7	0.000116782

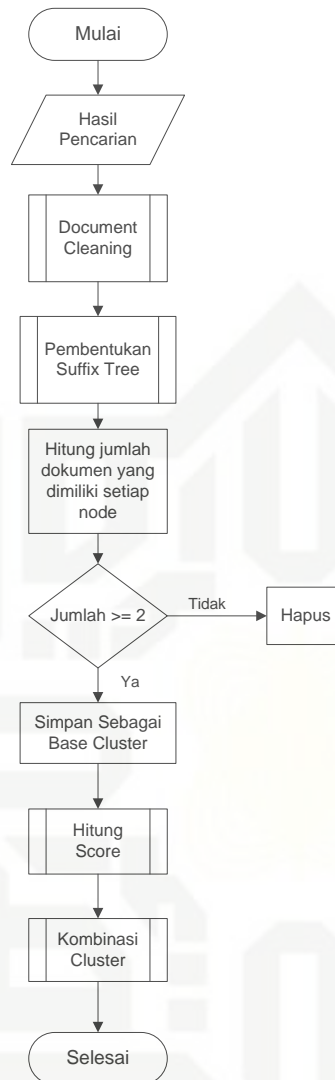
Dari hasil perhitungan *cosine similarity* tersebut diketahui bahwa Dokumen 13 (D13) memiliki nilai similaritas tertinggi kemudian diikuti D14, D9, D12, D3, D6, D4, D15, D1, D10.

4.4 Pengelompokan Hasil Pencarian (*Search Result Clustering*)

Setelah melakukan perhitungan untuk mendapatkan hasil pencarian yang relevan, proses selanjutnya yaitu melakukan pengelompokan terhadap hasil pencarian tersebut. Pengelompokan dilakukan untuk mempermudah menemukan hasil pencarian. Pengelompokan menggunakan metode *Suffix Tree Clustering*. Ada beberapa tahapan pada *Suffix Tree Clustering* yaitu *Document Cleaniing*, Pembentukan *Suffix Tree*, Identifikasi *Base Cluster* dan Kombinasi *Base Cluster*. Berikut *flowchart Suffix Tree Clustering*:

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 4.7 Flowchart Suffix Tree Clustering

Beikut diberikan contoh pengujian dengan menggunakan hasil pencarian dengan query “Recognition Using Backpropagation”. Hasil pencarian yang digunakan yaitu rangking 1-10.

Tabel 4.10 Contoh Kasus Clustering

Ranking	NO Dokumen	Judul
1	D13	Facial Expression Recognition By Using Fisherface Methode With Backpropagation Neural Network
2	D14	Online Handwritten Character Recognition Using an Optical Backpropagation Neural Network
3	D9	Impact of Ethnic Group on Human Emotion Recognition Using Backpropagation Neural Network
4	D12	Static Hand Gesture Recognition of Indonesian Sign Language System Based on Backpropagation Neural Networks

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.

5	D3	Multi-Digit Handwritten Sindhi Numerals Recognition using SOM Neural Network
6	D6	Spoof Detection for Finger-Vein Recognition System Using NIR Camera
7	D4	ALPHABET SIGN LANGUAGE RECOGNITION USING LEAP MOTION TECHNOLOGY AND RULE BASED BACKPROPAGATION-GENETIC ALGORITHM NEURAL NETWORK (RBBPGANN)
8	D15	Obstacle Recognition Based on Machine Learning for On-Chip LiDAR Sensors in a Cyber-Physical System
9	D1	A Robust Deep-Learning-Based Detector for Real-Time Tomato Plant Diseases and Pests Recognition
10	D10	An Adaptive Filtering Algorithm Based on Genetic Algorithm-Backpropagation Network

4.4.1 Document Cleaning

Document Cleaning merupakan proses pertama yang dilakukan untuk mendapatkan *base cluster* (*cluster* dasar) dari hasil pencarian. Pada proses ini hasil pencarian akan dibersihkan dari karakter yang tidak diperlukan seperti tanda baca dan penghapusan *stop-word* atau kata yang berfrekuensi banyak.

Tabel 4.11 Hasil *Document Cleaning*

Ranking	NO Dokumen	Judul
1	D13	facial expression recognition fisherface metode backpropagation neural network
2	D14	online handwritten character recognition optical backpropagation neural network
3	D9	impact ethnic group human emotion recognition backpropagation neural network
4	D12	static hand gesture recognition indonesian sign language system based backpropagation neural network
5	D3	multi digit handwritten sindhi numeral recognition som neural network
6	D6	spoof detection finger vein recognition system nir camera
7	D4	alphabet sign language recognition leap motion technology rule based backpropagation genetic algorithm neural network rbbpgan
8	D15	obstacle recognition based machine learning chip lidar sensor cyber physic system
9	D1	robust deep learn based detector real time tomato plant disease pests recognition
10	D10	adaptive filtering algorithm based genetic algorithm backpropagation network

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.

4.4.2 Pembentukan *Suffix Tree*

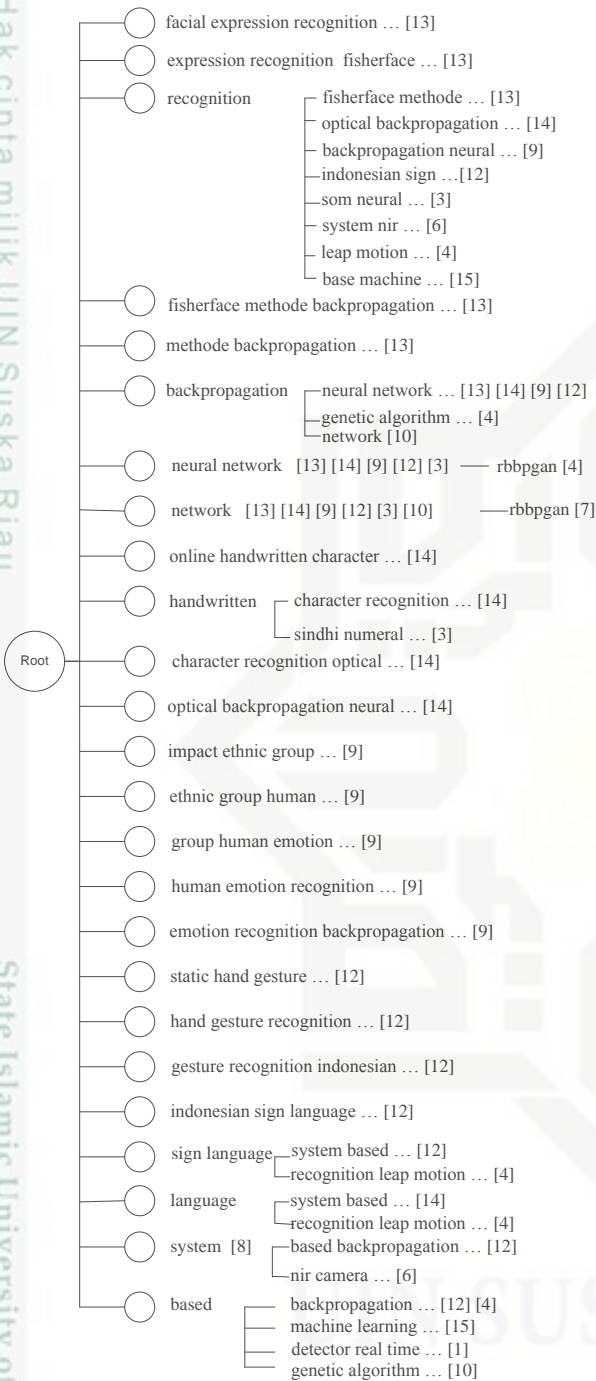
Proses pembentukan *Suffix Tree* merupakan proses terpenting pada *Suffix Tree Clustering*. Pembentukan *suffix tree* dilakukan dengan cara menemukan *share phares* antara dokumen.

Langkah-langkah pembentukan *suffix tree* ada pada Lampiran 1C. Berikut hasil *suffix tree* dari hasil pencarian. *Suffix* yang panjang akan diperpendek menggunakan notasi “...”



Hak Cipta Dilindungi Undang-Undang

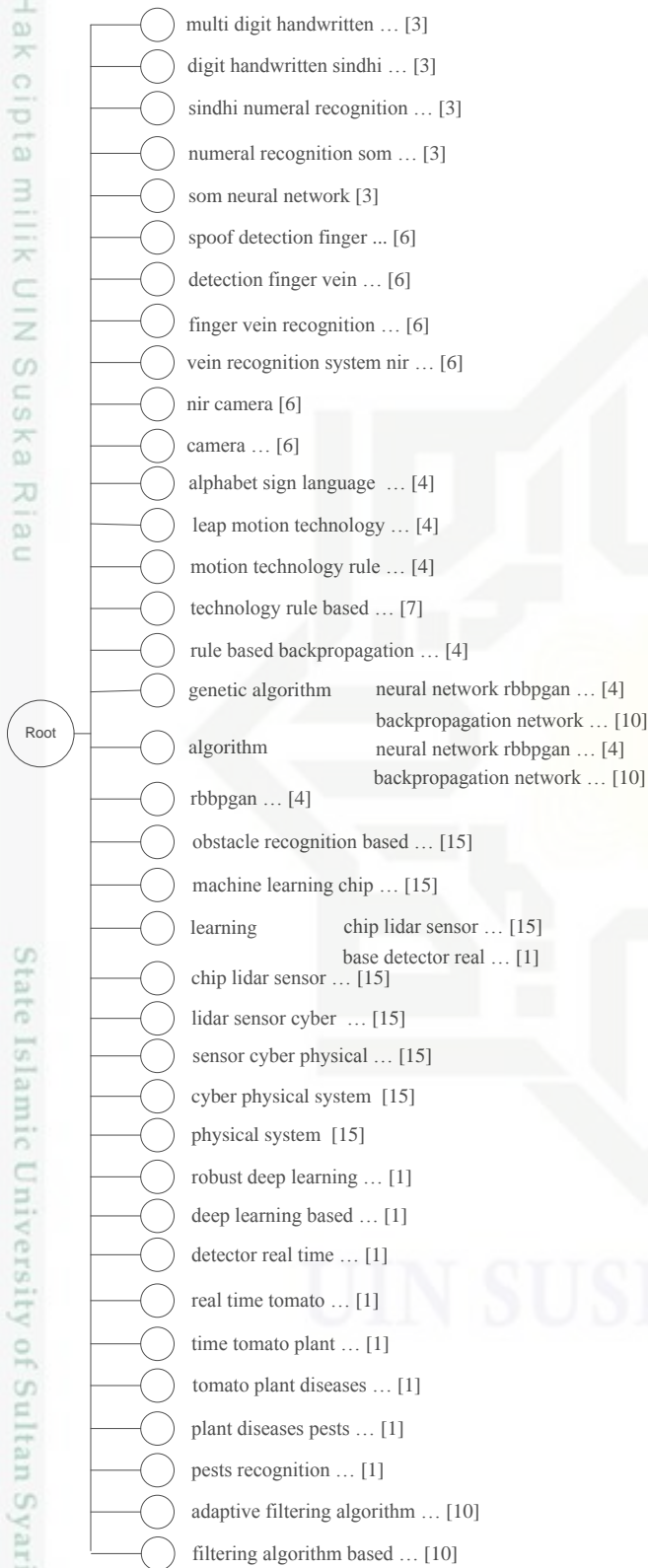
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan 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 4.8 Hasil Suffix Tree

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 4.9 Hasil Suffix Tree Lanjutan

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.

4.4.3 Identifikasi *Base Cluster*

Dari hasil *tree* yang terbentuk maka dipilih *node* yang termasuk *base cluster*. *Node* yang memiliki 2 dokumen atau lebih dipilih sebagai *base cluster*. Berikut *base cluster* yang dihasilkan dari *tree* yang terbentuk:

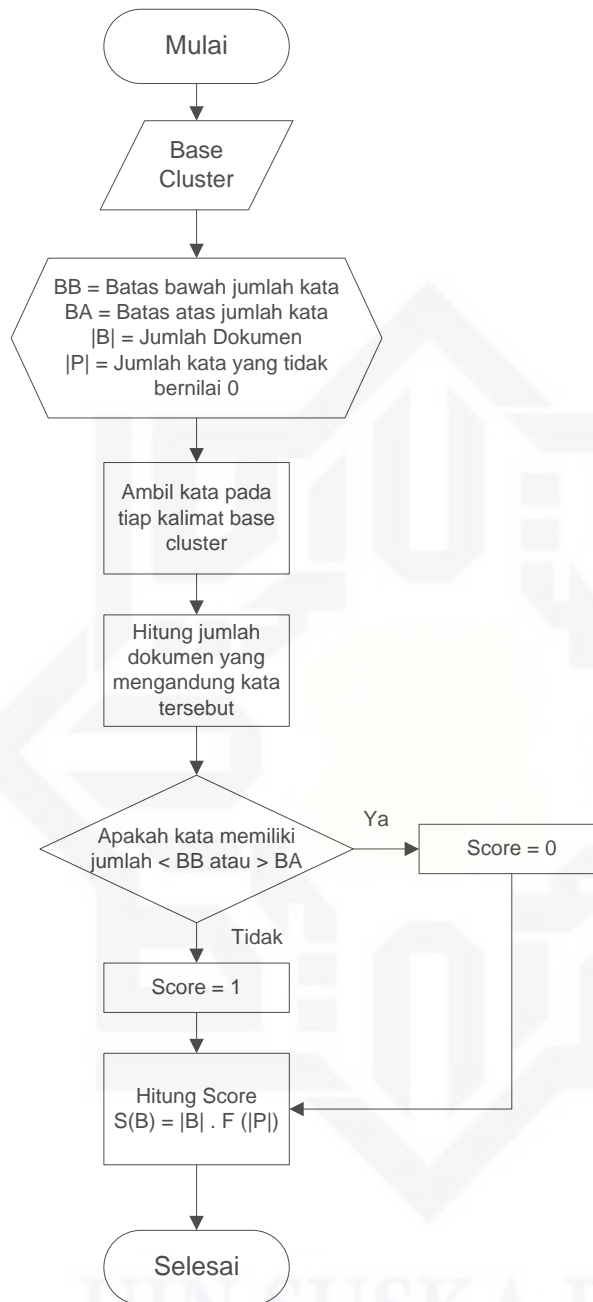
Tabel 4.12 *Base Cluster*

No	<i>Base Clutser</i>	Dokumen
1	Recognition	13,14,9,12,3,6,4,15,1
2	Backpropagation	13,14,9,12,4,10
3	Backpropagation Neural Network	13,14,9,12
4	Neural Network	13,14,9,12,3,4
5	Network	13,14,9,12,3,4,10
6	Handwritten	14,3
7	Sign Language	12,4
8	Language	12,4
9	Based Backpropagation	12,4
10	Based	12,4,15,1,10
11	Genetic Algorithm	4,10
12	Algorithm	4,10
13	Learning	15,1
14	System	12,6,15

Setelah mendapatkan *base cluster* proses selanjutnya yaitu menghitung *score base cluster* tersebut. Pada proses ini akan dihitung jumlah kemunculan setiap kata pada dokumen. Batas atas dan batas bawah jumlah kemunculan kata ditentukan oleh administrator. Jika kata-kata tersebut jumlahnya lebih kecil dari batas bawah atau lebih besar dari batas atas atau kata-kata tersebut termasuk *stop-word*, maka *score* kata tersebut 0. Lalu hitung jumlah dokumen pada *base cluster* tersebut. Setelah mendapatkan jumlah kata yang tidak memiliki *score* 0 dan jumlah dokumen maka hitung *score base cluster* dengan persamaan (2.6). Berikut *flowchart* menghitung *score base cluster*:

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 mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar 4.10 Menghitung Score Base Cluster

Contoh:

Base Cluster 3 Backpropagation Neural Network

Jumlah kata : 3 (*Backpropagation, Neural, Network*)

Jumlah Dokumen $|B|$: 4 (13, 14, 9, 12)

Batas Atas : 40% dari total semua dokumen

Batas Bawah : 10% dari total semua dokumen

Jumlah kemunculan kata *Backpropagation* pada keseluruhan dokumen yaitu 9. Jumlah kemunculan kata *Neural* pada keseluruhan dokumen yaitu 6. Jumlah kemunculan kata *Network* pada keseluruhan dokumen yaitu 7. Jumlah kata *Backpropagation*, *Neural* dan *Network* melewati batas atas yaitu 40% dari total dokumen atau 4 dari total 10 dokumen. Sehingga kata *Backpropagation*, *Neural* dan *Network* memiliki score 0. Maka jumlah *score base cluster*-nya yaitu:

$$s(B) = |B|.f(|P|)$$

$$= 4.0$$

$$= 0$$

Hasil perhitungan *score base cluster* secara keseluruhan pada tabel berikut

Tabel 4.13 Score Base Cluster

No	Base Clutser	Score
1	Recognition	0
2	Backpropagation	0
3	Backpropagation Neural Network	0
4	Neural Network	0
5	Network	0
6	Handwritten	2
7	Sign Language	4
8	Language	2
9	Based Backpropagation	0
10	Based	0
11	Genetic Algorithm	4
12	Algorithm	4
13	Learning	4
14	System	3

Setelah proses penghitungan *score base cluster* lalu dipilih *base cluster* yang tidak memiliki *score* 0 untuk dijadikan *cluster*. Berikut tabel hasil *cluster*

Tabel 4.14 Cluster

No	Base Clutser	Dokumen
1	Handwritten	14,3
2	Sign Language	12,4
3	Language	12,4
4	Genetic Algorithm	4,10
5	Algorithm	4,10
6	Learning	15,1
7	System	12,6,15

4.4.4 Kombinasi Base Cluster

Proses selanjutnya yaitu melakukan kombinasi antar *cluster*. Pada tahap ini dilakukan kombinasi terhadap *cluster* untuk mendapatkan hubungan antar *cluster*. Sebelum melakukan kombinasi antar *cluster*, terlebih dahulu dihitung nilai *similarity* antar *cluster*. Nilai *similarity* dihitung dengan persamaan (2.7) dan (2.8) dan nilai *threshold* yang digunakan yaitu 0.5. Berikut diberikan contoh perhitungan *similarity* antara *cluster* no. 2 yaitu *Sign Language* dan *cluster* no. 3 yaitu *language*.

- a. Jumlah dokumen pada *cluster* no. 2 *sign language* yaitu 2 (Dokumen 12 dan 4)
- b. Jumlah dokumen pada *cluster* no. 3 *language* yaitu 2 (Dokumen 12 dan 4)

Kedua *cluster* dioperasikan irisan, maka diperoleh dokumen yang berada di kedua simpul dokumen 12 dan 4 yaitu:

$$B_2 = 2$$

$$B_3 = 2$$

$$|B_2 \cap B_3| = 2$$

$$|B_2 \cap B_3| / |B_2| > 0.5 = \frac{2}{2} = 1 > 0.5$$

$$|B_2 \cap B_3| / |B_3| > 0.5 = \frac{2}{2} = 1 > 0.5$$

Dengan demikian kedua *cluster* memiliki nilai *similarity*, sehingga kedua *cluster* dapat berada dalam satu *cluster*. berikut rekap perhitungan nilai *similarity* antar *cluster*:

Tabel 4.15 Nilai Similarity antar cluster

Cluster	1	2	3	4	5	6	7
1	-	0	0	0	0	0	0
2	0	-	1	0.5	0.5	0	0.5
3	0	1	-	0.5	0.5	0	0.5
4	0	0.5	0.5	-	1	0	0.5
5	0	0.5	0.5	1	-	0	0
6	0	0	0	0	0	-	0.5
7	0	0.3	0.3	0	0	0	-

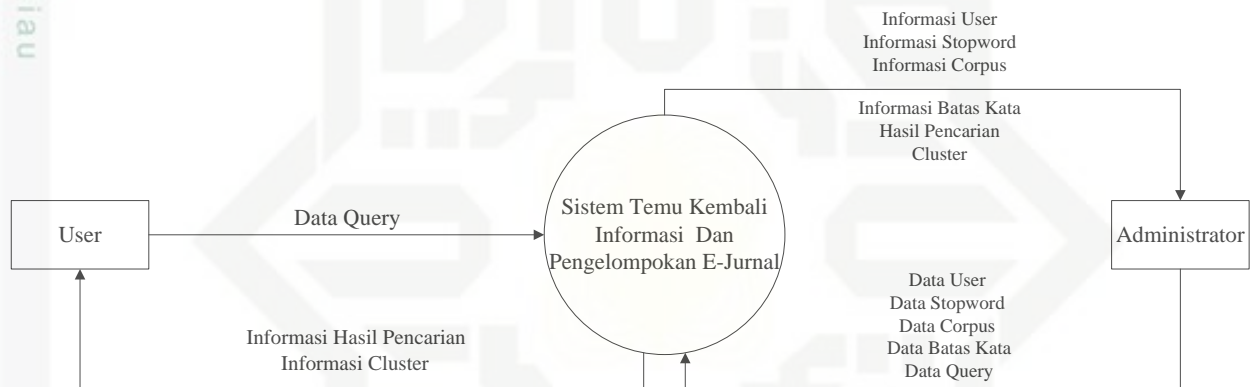
Pada tabel di atas dapat dilihat bahwa *cluster* yang memiliki nilai *similarity* lebih dari 0.5 yaitu *cluster* no. 2 dan 3 dan *cluster* no. 4 dan 5. *Cluster-cluster* tersebut digabungkan menjadi satu *cluster* dan ditampilkan dengan nama *cluster* yang memiliki *score* tertinggi .

4.5 Perancangan Sistem

Perancangan merupakan tahapan untuk menggambarkan dan memodelkan sistem yang akan dibuat. Pada penelitian ini perancangan sistem meliputi *Context Diagram*, *Data Flow Diagram* dan Rancangan Basis Data.

4.5.1 Context Diagram

Context Diagram adalah model proses yang menggambarkan ruang lingkup suatu sistem. *Context Diagram* hanya mempunyai satu proses (Whitten & Bentley, 2007). Berikut adalah *Context Diagram* Sistem Temu Kembali Informasi dan Pengelompokan e-Jurnal:



Gambar 4.11 Context Diagram Sistem Temu Kembali Informasi dan Pengelompokan E-Jurnal

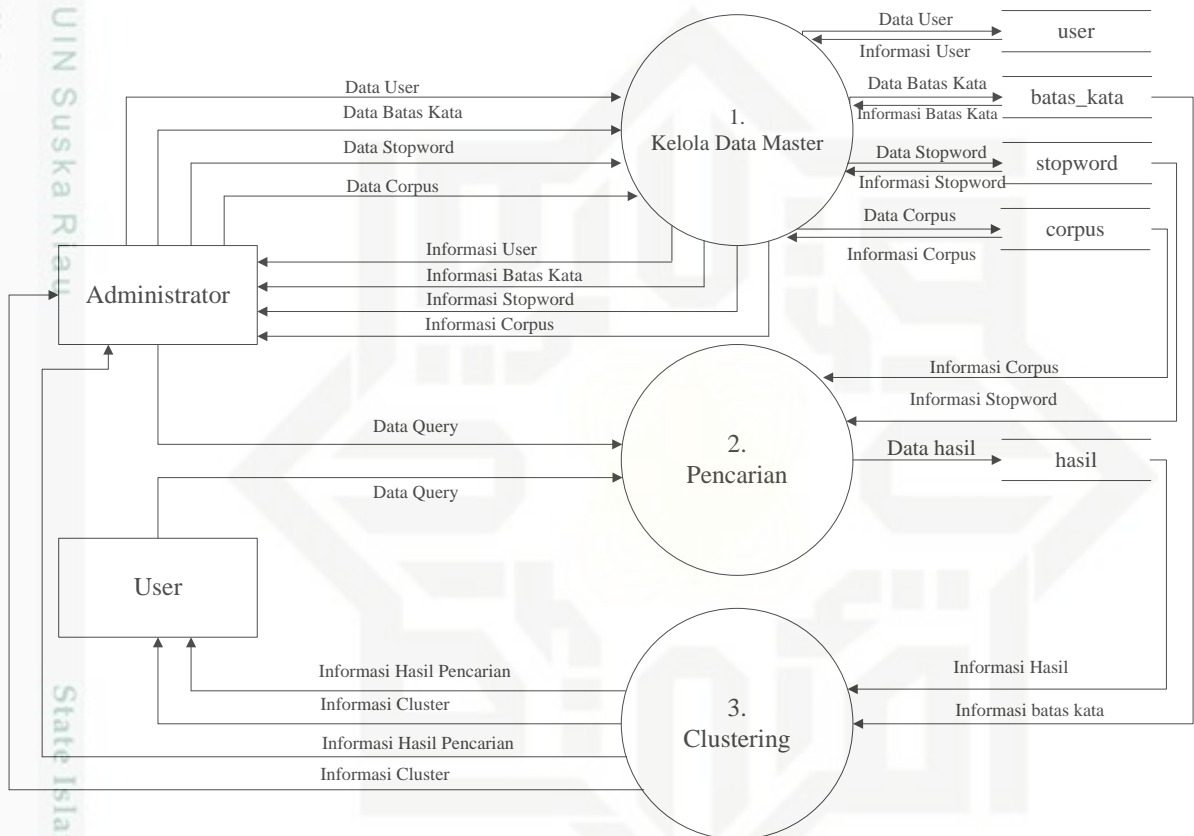
Berikut deskripsi *Context Diagram* Sistem Temu Kembali Informasi dan Pengelompokan e-Jurnal:

Tabel 4.16 Deskripsi Context Diagram

NO	Entitas	Input	Output	Keterangan
1	Administrator	User, Stopword, Corpus, Batas Kata dan Query	Informasi User, Informasi Stopword, Informasi Corpus, Informasi Batas Kata, Hasil Pencarian dan Cluster	Administrator dapat mengelola Data Master dan melakukan pencarian pada sistem
2	User	Query	Hasil Pencarian dan Cluster	User hanya bisa melakukan pencarian pada sistem

4.5.2 Data Flow Diagram (DFD) Level 1

Data Flow Diagram (DFD) yaitu model proses untuk menggambarkan aliran data suatu sistem dan proses yang dilakukan oleh sistem (Whitten & Bentley, 2007). Berikut adalah Data Flow Diagram (DFD) Level 1 Sistem Temu Kembali Informasi dan Pengelompokan e-Jurnal:



Gambar 4.12 Data Flow Diagram (DFD) Level 1

Berikut deskripsi Data Flow Diagram Level 1 Proses 1 Sistem Temu Kembali Informasi dan Pengelompokan e-Jurnal:

Tabel 4.17 Deskripsi Data Flow Diagram Level 1 Proses 1

No. Proses	1
Nama Proses	Kelola Data Master
Deskripsi	Administrator dapat mengelola User, Corpus, Stopword dan Batas Kata
Input	Data User (username dan password) Data Corpus (Title, Author, Abstact, Affiliation, Publisher, Year, ISSN, DOI, Keywords, Link, Subject) Data Stopword (Nama Stopword) Batas Kata (Batas Atas, Batas Bawah)
Output	Data User, Data Corpus, Data Stopword dan Data Batas Kata

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.

Berikut deskripsi *Data Flow Diagram Level 1* Proses 2 Sistem Temu Kembali Informasi dan Pengelompokan e-Jurnal:

Tabel 4.18 Deskripsi *Data Flow Diagram Level 1* Proses 2

No. Proses	2
Nama Proses	Pencarian
Deskripsi	User dan Administrator dapat melakukan pencarian
Input	<i>Query</i>
Output	Hasil Pencarian

Berikut deskripsi *Data Flow Diagram Level 1* Proses 3 Sistem Temu Kembali Informasi dan Pengelompokan e-Jurnal:

Tabel 4.19 Deskripsi *Data Flow Diagram Level 1* Proses 3

No. Proses	3
Nama Proses	<i>Clustering</i>
Deskripsi	Hasil Pencarian yang dilakukan user akan di kelompokkan
Input	Hasil Pencarian dan Batas Kata
Output	<i>Cluster</i> Hasil Pencarian

4.5.3 Perancangan Basis Data

Rancangan Basis Data merupakan tahapan analisa tabel yang dibutuhkan pada pembuatan sistem. Tabel 4.20 berikut adalah deskripsi tabel yang akan dirancang:

Tabel 4.20 deskripsi tabel yang dirancang pada *database*

No	Nama Tabel	Deksripsi	Attribute	Primary Key
1	User	Tabel penyimpanan Data Username dan Password Administrator	- Id_User - Username - Password	id_User
2	Corpus	Tabel penyimpanan Data Corpus	- Id_corpus - Title - Year - Abstract - Author - Affilation - Subjek - Keyword - Link - Doi - Publisher - Issn	Id_corpus

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.

No	Nama Tabel	Deksripsi	Attribute	Primary Key
3	Stopword	Tabel penyimpanan Data Stopword	- Id_Stopword - Stopword	Id_Stopword
4	Hasil	Tabel penyimpanan hasil pencarian yang sudah di preprocessing untuk proses <i>clustering</i>	- Id - hasil	id
5	batas_kata	Tabel penyimpanan Data Batas Kata	- Id_Batas_Kata - Batas_Atas - Batas_Bawah	Id_Batas_Kata

4.6 Perancangan Antarmuka

Perancangan antarmuka merupakan rancangan tampilan yang akan digunakan pada sistem yang dibangun. Berikut rancangan antarmuka sistem temu kembali dan pengelompokan e-Jurnal.

a. Halaman Depan Sistem

Berikut rancangan antarmuka halaman depan sistem

STKI E-Jurnal
Sign In

Gambar 4.13 rancangan antarmuka halaman depan sistem

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.

b. Halaman Hasil Pencarian

Berikut rancangan antarmuka halaman hasil pencarian:

STKI E-Jurnal		Sign In
<input style="width: 80%; border: 1px solid black;" type="text" value="Input Query"/>		<input style="border: 1px solid black; border-radius: 15px; padding: 2px 10px;" type="button" value="Search"/>
Cluster 1	Hasil Pencarian 1	
Cluster 2	Hasil Pencarian 2	
Cluster 3	Hasil Pencarian 3	

Gambar 4.14 rancangan antarmuka halaman hasil pencarian

c. Halaman Jurnal

Berikut rancangan antar muka halaman jurnal

STKI E-Jurnal		Sign In
Judul Jurnal		
Author		
Affiliation		
Tahun : Year		
Publisher: Publisher		
ISSN : No ISSN		
Doi : No DOI		
Keyword: Keyword1, Keyword 2, Keyword 3, Keyword 4, Keyword 5		
Subjek : Subjek		
Abstrak Link		
Isi Abstrak		

Gambar 4.15 rancangan antarmuka halaman halaman jurnal