

# PENGUNAAN METODE *COSINESIMILARITY* PADA SISTEM PENGELOMPOKAN KERJA PRAKTEK, TUGAS AKHIR DAN SKRIPSI

Randy Agung Wibowo (agung.wbowo92@gmail.com)  
 Didik Nugroho (masdidiknugroho@gmail.com)  
 Bebas Widada(bbswdd@gmail.com)

## ABSTRAK

Pengumpulan laporan kerja praktik, tugas akhir dan skripsi mahasiswa di perpustakaan STMIK Sinar Nusantara Surakarta dikelompokkan berdasarkan jurusan yang ada di STMIK Sinar Nusantara. Laporan tersebut tidak dikelompokkan berdasarkan konsentrasi ilmu teknik informatika yang ada seperti sistem informasi, jaringan komputer, hardware dan multimedia. Hal ini akan menyulitkan mahasiswa lainnya jika mencari laporan berdasarkan konsentrasi ilmu teknik informatika tertentu. Sistem pengelompokan otomatis menggunakan bahasa pemrograman PHP dan menggunakan DBMS MySQL. Dokumen yang dikategorikan akan melalui proses tokenisasi dan filterisasi sehingga menghasilkan sebuah term yang nantinya akan dihitung jumlah kata yang mirip dalam satu dokumen. Kemudian jumlah kata tersebut akan digunakan dalam perhitungan dengan menggunakan metode *cosinesimilarity*. Berdasarkan penelitian yang telah dilakukan diperoleh kesimpulan bahwa penerapan metode *cosinesimilarity* pada sistem pengelompokan otomatis menghasilkan sebuah sistem yang dapat mengelompokkan dokumen kedalam masing-masing kategori. Penentuan kategori dari sebuah dokumen yang telah dilakukan proses perhitungan adalah dengan mengambil nilai *cosinesimilarity* yang paling besar terhadap pada sebuah kategori. Berdasarkan hasil sampling yang dilakukan dapat diketahui bahwa faktor yang mempengaruhi dokumen masuk kedalam sebuah kategori adalah kata-kata penyusun dokumen tersebut.

**KATA KUNCI** : pengelompokan dokumen, *cosinesimilarity*, *term*, tokenisasi, filterisasi

## I. PENDAHULUAN

### 1.1 Latar Belakang

Karya ilmiah adalah laporan tertulis dan diterbitkan yang memaparkan hasil penelitian atau pengkajian yang telah dilakukan oleh seseorang atau sebuah tim dengan memenuhi kaidah dan etika keilmuan yang dikukuhkan dan ditaati oleh masyarakat keilmuan. Ada berbagai jenis karya ilmiah, antara lain laporan penelitian, makalah seminar atau simposium, dan artikel jurnal yang pada dasarnya kesemuanya itu merupakan produk dari kegiatan ilmuwan.

Setiap perguruan tinggi memiliki banyak karya ilmiah khususnya dengan jenis laporan penelitian. Seperti perguruan tinggi yang lain Sekolah Tinggi Manajemen Informatika Dan Komputer Sinar Nusantara (STMIK Sinus) Surakarta memiliki beberapa karya ilmiah, dengan berbagai berjenis laporan penelitian. Laporan penelitian tersebut terdiri dari laporan kerja praktik, laporan laporan dan laporan skripsi. Laporan-laporan tersebut dikumpulkan setelah mahasiswa selesai melakukan

pendadaran dan dikumpulkan pada perpustakaan.

Pengumpulan laporan kerja praktik, tugas akhir dan skripsi mahasiswa di perpustakaan STMIK Sinar Nusantara Surakarta dikelompokkan berdasarkan jurusan yang ada di STMIK Sinar Nusantara. Laporan tersebut tidak dikelompokkan berdasarkan konsentrasi ilmu teknik informatika yang ada seperti sistem informasi, jaringan komputer, hardware dan multimedia. Hal ini akan menyulitkan mahasiswa lainnya jika mencari laporan berdasarkan konsentrasi ilmu teknik informatika tertentu.

Dari permasalahan diatas penulis akan melakukan penelitian tentang sebuah sistem pengelompokkan laporan penelitian secara otomatis dengan menggunakan metode cosine similarity sehingga dapat membantu mahasiswa dalam mencari laporan yang memiliki tingkat kesamaan yang tinggi terhadap suatu konsentrasi ilmu informatika yang ada. Metode cosine similarity merupakan salah satu metode yang populer dalam mengukur kedekatan teks. Metode ini akan menghitung nilai cosinus antara dua sudut

vector. Apabila nilai cosinus yang dihitung mendekati angka 1 maka teks yang dihitung akan menunjukkan bahwa dokumen tersebut semakin mirip dengan query yang telah ditentukan..

## 1.2 Tujuan

Tujuan dari penelitian ini adalah menerapkan metode *cosinesimilarity* pada sistem pengelompokan laporan kerja praktek, tugas akhir dan skripsi jurusan S1 Sistem Informasi, S1 Teknik Informatika, D3 Manajemen Informatika, D3 Komputerisasi Akuntansi dan D3 Teknik Informatika STMIK Sinar Nusantara Surakarta.

## II. METODE PENELITIAN

### 2.1 Data

#### • Data Primer

Data primer diperoleh dari perpustakaan STMIK Sinus Surakarta. Dari perpustakaan STMIK Sinus Surakarta diperoleh data abstraksi laporan mahasiswa dimana laporan tersebut termasuk laporan kerja praktik, laporan tugas akhir dan laporan skripsi. Data tersebut merupakan data utama yang akan dikelola.

#### • Data Sekunder

Data sekunder yang diperlukan adalah data daftar kata-kata yang kurang memiliki makna dan data kata-kata atau istilah yang mencerminkan suatu kategori tertentu. Daftar kata-kata yang tidak memiliki makna atau biasa disebut dengan stoplist diperoleh dari penelitian yang dilakukan oleh Tara (2003). Sedangkan data istilah yang mencerminkan suatu kategori tertentu diperoleh dari penelitian yang dilakukan oleh Purwaningrum<sup>[1]</sup>.

### 2.2 Alur Penelitian

Alur jalannya penelitian dibuatnya system pengelompokan adalah sebagai berikut:

#### • Persiapan

Tahap persiapan meliputi observasi perpustakaan, pengumpulan data dan observasi metode.

#### 1. Observasi perpustakaan

Pada tahap ini dilakukan identifikasi mengenai proses bisnis pengumpulan laporan kerja praktik, laporan tugas akhir dan laporan skripsi. Selain proses bisnis, diidentifikasi pula berkas-berkas yang akan dikumpulkan di perpustakaan.

#### 2. Pengumpulan Data

Pengumpulan data terbagi menjadi dua metode. Metode antara lain adalah metode wawancara dan metode studi pustaka.

#### a. Wawancara

Narasumber dalam wawancara adalah petugas perpustakaan guna melengkapi data penelitian. Wawancara yang dilakukan mengenai pengelompokan secara manual yang dilakukan oleh petugas perpustakaan terhadap laporan kerja praktik, laporan tugas akhir, dan laporan skripsi.

#### b. Studi Pustaka

Studi pustaka dilakukan guna melengkapi data sekunder yang berupa daftar istilah yang mencerminkan tiap kategori. Daftar ini diperoleh dengan mengumpulkan laporan-laporan yang sudah diketahui kelompoknya kemudian mencari kata-kata khusus yang sering muncul pada kelompok tersebut.

### 3. Metode Pencarian

Metode yang dapat digunakan dalam system pengelompokan merupakan metode pembobotan kata. Pada kasus ini akan diambil sebagian rumus dari metode *Term Frequency-Inverse Document Frequency*, dimana akan digunakan nilai *tf* yang akan dihitung lagi dengan menggunakan metode *cosine similarity*.

#### • Pelaksanaan

##### 1. Perancangan Basis Data

Tahap perancangan basis data dibagi 3 tahapan, yaitu:

- Membuat *Diagram Entity-Relationship (ERD)*.
- Membuat Skema Diagram.
- Membuat Struktur *Database*.

##### 2. Perancangan Sistem

Tahap perancangan sistem dibagi 2, yaitu :

- Membuat *Context Diagram*.
- Membuat *Data Flow Diagram*.

##### 3. Implementasi Metode

##### 4. Perancangan Desain *Interface*

#### • Implementasi

Pada tahap ini dibuat system dengan mengimplementasikan rancangan *database*, rancangan sistem, perhitungan metode dan rancangan *interface* yang telah dibuat pada tahap sebelumnya. Hasil dari pengimplementasian berupa terbentuknya system pengelompokan yang dapat

mengelompokkan laporan sesuai dengan isi dari laporan tersebut.

Pada tahap uji coba akan dilakukan testing sistem dengan menggunakan 100 data laporan. Data laporan tersebut terdiri dari laporan kerja praktik, laporan tugas akhir dan laporan skripsi.

## 2.3 Alat dan Bahan

### a. Alat dan Bahan

Perangkat keras yang dibutuhkan dalam pembuatan sistem pengelompokan ini adalah sebagai berikut:

1. *Processor Intel Pentium P6200*
2. *Harddisk minimal 70 GB*
3. *RAM DDR2 minimal 1 GB*
4. *Keyboard dan Mouse*

Perangkat lunak yang digunakan dalam pembuatan sistem pengelompokan adalah sebagai berikut:

1. Sistem Operasi *Microsoft Windows 7 Ultimate*
2. *Notepad++* yang digunakan untuk membuat program
3. *Dreamweaver CS5* yang digunakan untuk mengatur tampilan sistem
4. *XAMPP v1.7.7* yang terdiri dari *MySQL* dan *Apache*
5. *Balsamiq Mockups* untuk membuat desain *interface*

Bahan yang digunakan adalah:

1. Data laporan yang berupa nim mahasiswa, nama mahasiswa, tahun lulus, judul laporan, abstrak laporan, dan keyword pada abstrak
2. Data keyword yaitu daftar istilah yang mencerminkan masing-masing kategori yaitu kategori system informasi, jaringan, *hardware*, dan multimedia.

## 2.4 Pengujian

Pengujian yang penulis lakukan dengan melalui kuisioner:

1. Membagikan lembar kuisioner kepada pemakai
2. Mengumpulkan kembali data yang kuisioner yang telah diisi oleh pemakai
3. Penulis melakukan olah data kuisioner yang telah terkumpul.

## III. TINJAUAN PUSTAKA

### 3.1. Text Processing

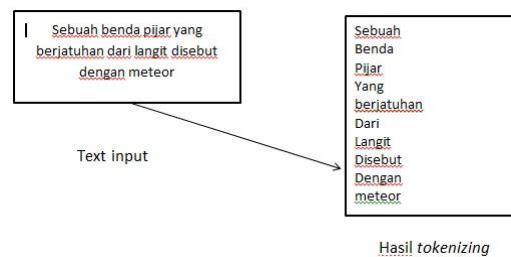
*Text* atau biasa disebut dengan korpus merupakan kumpulan dokumen yang dibaca

oleh mesin. Pengolahan teks atau *text processing* merupakan suatu cara yang digunakan untuk mengenali karakteristik dari sampel teks.

Tahap-tahap umumpemrosesan awal (*pre-processing*) yang dilakukan dalam *text processing* pada dokumen terdiri dari [1]:

#### 1. Tokenizing

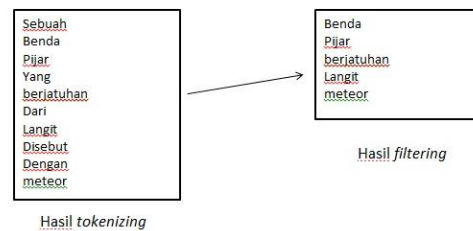
*Tokenizing* merupakan tahap pemotongan *string input* berdasarkan tiap kata yang menyusunnya. Contoh dari tahap ini adalah sebagai berikut:



Gambar 1. Proses *Tokenizing*

#### 2. Filtering

*Filtering* merupakan tahap pengambilan kata-kata penting hasil token. Bisa menggunakan algoritma *stoplist* (membuang kata yang kurang penting) atau *wordlist* (menyimpan kata penting). *Stoplist / stopword* adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan *bag-of-words*. Contoh *stopwords* adalah “yang”, “dan”, “di”, “dari” dan seterusnya[2]. Contoh dari tahapan ini adalah sebagai berikut[3]:



Gambar 2. Proses *Filtering*

#### 3.2. Term Frequency

*Term Frequency (tf) factor*, faktor yang menentukan bobot *term* pada suatu dokumen berdasarkan jumlah kemunculannya dalam dokumen tersebut. Nilai jumlah kemunculan suatu kata (*term frequency*) diperhitungkan dalam pemberian bobot terhadap suatu kata. Semakin besar jumlah kemunculan suatu *term (tf)* tinggi dalam dokumen, semakin besar pula bobotnya dalam dokumen atau akan memberikan nilai kesesuaian yang semakin besar.[4]

Pada *Term Frequency (tf)*, terdapat beberapa jenis formula yang dapat digunakan yaitu[5] :

1. *tf* biner (*binary tf*), hanya memperhatikan apakah suatu kata ada atau tidak dalam dokumen, jika ada diberi nilai satu, jika tidak diberi nilai nol.
2. *tf* murni (*raw tf*), nilai *tf* diberikan berdasarkan jumlah kemunculan suatu kata di dokumen. Contohnya, jika muncul lima kali maka kata tersebut akan bernilai lima.
3. *tf* logaritmik, hal ini untuk menghindari dominansi dokumen yang mengandung sedikit kata dalam *query*, namun mempunyai frekuensi yang tinggi.

$$tf = 1 + \log(tf) \dots \dots \dots (1)$$

4. *tf* normalisasi, menggunakan perbandingan antara frekuensi sebuah kata dengan jumlah keseluruhan kata pada dokumen.

$$tf = 0.5 + 0.5 \frac{tf}{\max tf} \dots \dots \dots (2)$$

## 2.2. Pengukuran *Similaritas* dengan *Cosine Similarity*

Model ruang vektor dan pembobotan *tf-idf* digunakan untuk merepresentasikan nilai numerik dokumen sehingga kemudian dapat dihitung kedekatan antar dokumen. Kemiripan antar dokumen dihitung menggunakan suatu fungsi ukuran kemiripan (*similaritymeasure*). Semakin besar hasil fungsi *similarity*, maka kedua objek yang dievaluasi semakin mirip, demikian pula sebaliknya. Ukuran ini memungkinkan perankingan dokumen sesuai dengan kemiripan (relevansi)nya terhadap *query*. Kualitas hasil dari dokumen yang didapatkan sangat tergantung pada fungsi *similarity* yang digunakan.[6]

Jika terdapat dua vektor dokumen *d* dan *q* serta *term t* diperoleh dari dokumen maka nilai cosinus didefinisikan sebagai:

$$similarity(\vec{d}_j, \vec{q}) = \frac{\vec{d}_j \cdot \vec{q}}{|\vec{d}_j| \cdot |\vec{q}|} \dots \dots \dots (3)$$

keterangan:

$\vec{d}_j$  = vektor dokumen ke-*j*(*tf d*)

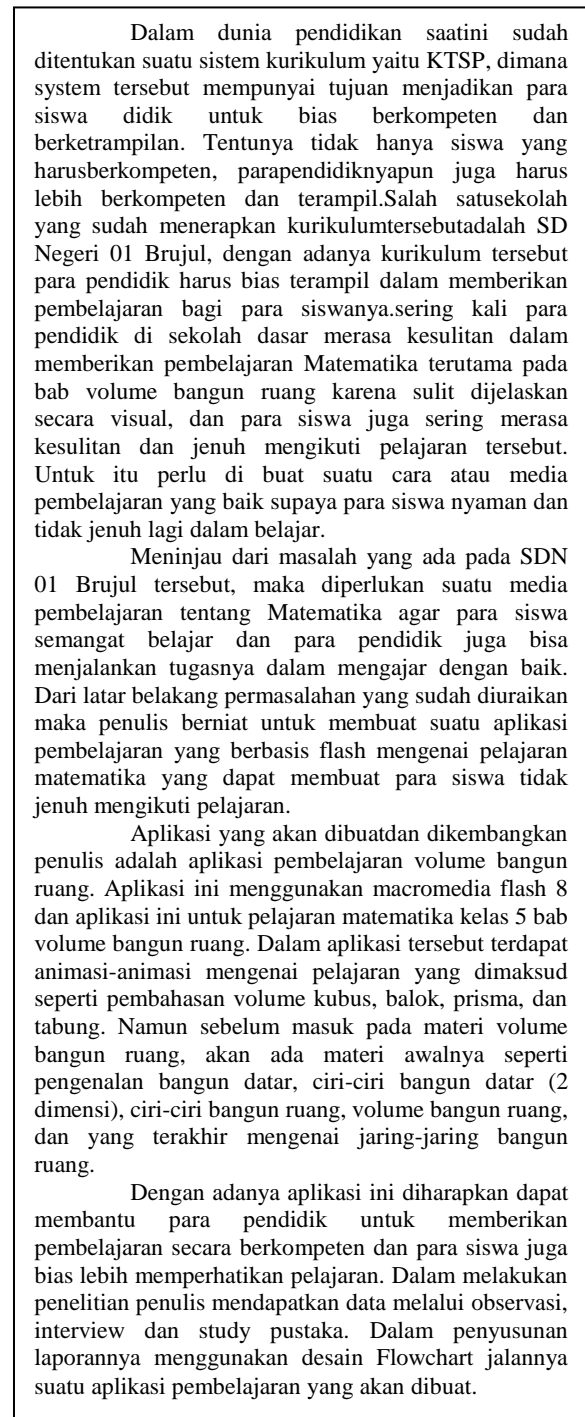
$\vec{q}$  = vektor *q*(*tf q*)

## IV. HASIL DAN PEMBAHASAN

### 4.1 PEMBAHASAN

Metode perhitungan *cosine similarity* dilakukan secara manual dan secara otomatis menggunakan sistem pengelompokan. Untuk melakukan pengujian, berikut merupakan

sebuah contoh data abstraksi yang ditunjukkan pada Gambar 2.



Gambar 2. Contoh Abstrak

Berdasarkan contoh abstrak pada Gambar 2, dapat dilakukan proses *preprocessing* dan penghitungan nilai *term frequency*. Dengan menggunakan perhitungan secara manual abstraksi diatas akan dipisahkan setiap katanya menjadi sebuah *term*. Kemudian *term* tersebut akan melalui proses *filtering* dengan menghapus kata-kata yang tidak memiliki makna. Setelah itu akan

dicari jumlah kata yang sama atau *termfrequency*. Hasil dari perhitungan secara manual disajikan pada Tabel 2.

**Tabel 2.** Hasil Proses *Preprocessing*

Term	TF	Term	TF
bangun	9	tabung	1
siswa	8	masuk	1
aplikasi	8	desain	1
pembelajaran	8	balok	1
ruang	7	kubus	1
pelajaran	6	jalannya	1
volume	6	kelas	1
pendidik	4	flowchart	1
matematika	4	pengenalan	1
berkompeten	4	pembahasan	1
jenuh	3	r	1
penulis	3	penyusunan	1
kurikulum	3	memperhatikan	1
bab	2	penelitian	1
flash	2	macromedia	1
animasi	2	jaring-jaring	1
seperti	2	diharapkan	1
ciri-ciri	2	data	1
belajar	2	dimensi	1
media	2	pustaka	1
materi	2	membantu	1
mengikuti	2	study	1
datar	2	interview	1
kesulitan	2	observasi	1
sistem	2	laporannya	1
terampil	2	semangat	1
brujul	2	satu	1
sekolah	2	salah	1
prisma	1	pendidiknyapun	1
n	1	dasar	1
menerapkan	1	latar	1
sd	1	mengajar	1
nya	1	permasalahan	1
negeri	1	diuraikan	1
berketrampilan	1	berbasis	1
didik	1	berniat	1
ditentukan	1	tugasnya	1
pendidikan	1	menjalankan	1
ktsp	1	visual	1
dimana	1	sulit	1
menjadikan	1	nyaman	1
tujuan	1	meninjau	1
kali	1	dunia	1
dikembangkan	1	sdn	1

Tabel diatas merupakan hasil dari proses *tokenizing*, *filtering* dan pencarian nilai *termfrequency*. Berdasarkan tabel tersebut dapat dicari nilai *cosinesimilarity* untuk masing-masing kategori sebagai berikut:

**a. Kategori 1 (Sistem Informasi)**

*Query* atau *keyword* yang mencerminkan kategori 1 berjumlah 26 *query* dan masing-masing nilai *termfrequencyquery* (tfq)

adalah 1, maka perhitungan nilai *cosinesimilarity* adalah:

$$\begin{aligned} \cos(\vec{tfd} \cdot \vec{tfq})_1 &= \frac{tfd_1tfq_1 + tfd_2tfq_2 + \dots + tfd_n tfq_n}{\sqrt{tfd_1^2 + tfd_2^2 + tfd_3^2} \sqrt{tfq_1^2 + tfq_2^2 + \dots + tfq_n^2}} \\ \cos(\vec{tfd} \cdot \vec{tfq})_1 &= \frac{2.1 + 1.1 + 8.1}{\sqrt{2^2 + 1^2 + 8^2} \sqrt{26 \cdot (1^2)}} \\ \cos(\vec{tfd} \cdot \vec{tfq})_1 &= \frac{11}{\sqrt{69} \sqrt{26}} \\ \cos(\vec{tfd} \cdot \vec{tfq})_1 &= \frac{11}{\sqrt{1794}} = 0.259 \end{aligned}$$

**b. Kategori 2 (Jaringan)**

*Query* atau *keyword* yang mencerminkan kategori 2 berjumlah 27 *query* dan masing-masing nilai *term frequency query* (tfq) adalah 1, maka perhitungan nilai *cosine similarity* adalah:

$$\begin{aligned} \cos(\vec{tfd} \cdot \vec{tfq})_2 &= \frac{tfd_1tfq_1 + tfd_2tfq_2 + \dots + tfd_n tfq_n}{\sqrt{tfd_1^2 + tfd_2^2 + tfd_3^2} \sqrt{tfq_1^2 + tfq_2^2 + \dots + tfq_n^2}} \\ \cos(\vec{tfd} \cdot \vec{tfq})_2 &= \frac{0}{\sqrt{0} \sqrt{27 \cdot (1^2)}} \\ \cos(\vec{tfd} \cdot \vec{tfq})_2 &= \frac{0}{\sqrt{0} \sqrt{27}} \\ \cos(\vec{tfd} \cdot \vec{tfq})_2 &= 0 \end{aligned}$$

**c. Kategori 3 (Hardware)**

*Query* atau *keyword* yang mencerminkan kategori 3 berjumlah 14 *query* dan masing-masing nilai *term frequency query* (tfq) adalah 1, maka perhitungan nilai *cosine similarity* adalah:

$$\begin{aligned} \cos(\vec{tfd} \cdot \vec{tfq})_3 &= \frac{tfd_1tfq_1 + tfd_2tfq_2 + \dots + tfd_n tfq_n}{\sqrt{tfd_1^2 + tfd_2^2 + tfd_3^2} \sqrt{tfq_1^2 + tfq_2^2 + \dots + tfq_n^2}} \\ \cos(\vec{tfd} \cdot \vec{tfq})_3 &= \frac{0}{\sqrt{0} \sqrt{14 \cdot (1^2)}} \\ \cos(\vec{tfd} \cdot \vec{tfq})_3 &= \frac{0}{\sqrt{0} \sqrt{14}} \\ \cos(\vec{tfd} \cdot \vec{tfq})_3 &= 0 \end{aligned}$$

#### d. Kategori 4 (Multimedia)

Query atau keyword yang mencerminkan kategori 4 berjumlah 17 query dan masing-masing nilai term frequency query (tfq) adalah 1, maka perhitungan nilai cosine similarity adalah:

$$\begin{aligned} & \cos(\overrightarrow{tfd} \cdot \overrightarrow{tfq})_4 \\ &= \frac{tfd_1 tfq_1 + tfd_2 tfq_2 + \dots + tfd_n tfq_n}{\sqrt{tfd_1^2 + tfd_2^2 + tfd_3^2} \sqrt{tfq_1^2 + tfq_2^2 + \dots + tfq_n^2}} \\ & \cos(\overrightarrow{tfd} \cdot \overrightarrow{tfq})_4 \\ &= \frac{2.1 + 2.1 + 1.1 + 8.1 + 2.1 + 1.1}{\sqrt{2^2 + 2^2 + 1^2 + 8^2 + 2^2 + 1^2} \sqrt{17 \cdot (1^2)}} \\ & \cos(\overrightarrow{tfd} \cdot \overrightarrow{tfq})_4 = \frac{16}{\sqrt{78} \sqrt{17}} \\ & \cos(\overrightarrow{tfd} \cdot \overrightarrow{tfq})_4 = \frac{16}{\sqrt{1326}} = 0.439 \end{aligned}$$

Dari hasil perhitungan nilai cosinesimilarity masing-masing kategori diambil nilai cosinesimilarity yang paling besar. Pada perhitungan manual diatas menunjukkan bahwa abstraksi tersebut masuk kedalam kategori 4 yaitu multimedia.

Pada sistem pengelompokan proses tokenizing, filtering dan pencarian term frequency menghasilkan nilai yang sama dengan penghitungan manual. Berikut merupakan hasil pencarian nilai cosinesimilarity dengan menggunakan sistem pengelompokan :

#### Proses Cosine Similarity

kategori 1:

```
sistem ---> tfd=2 ---> tfq=1
data ---> tfd=1 ---> tfq=1
aplikasi ---> tfd=8 ---> tfq=1

jumlah (tfd * tfq) = 11
jumlah (tfd * tfd) = 69
jumlah (tfq * tfq) = 26
akar (tfd * tfd) = 8.3066238629181
akar (tfq * tfq) = 5.0990195135928
nilai cosine = jumlah (tfd * tfq) / (akar (tfd * tfd) * akar (tfq * tfq)) = 0.25970569055744
```

kategori 2:

```
jumlah (tfd * tfq) = 0
jumlah (tfd * tfd) = 0
jumlah (tfq * tfq) = 27
akar (tfd * tfd) = 0
akar (tfq * tfq) = 5.1961524227066
nilai cosine = jumlah (tfd * tfq) / (akar (tfd * tfd) * akar (tfq * tfq)) =
```

**Gambar 3.** Nilai cosine similarity kategori 1 dan 2

kategori 3:

```
jumlah (tfd * tfq) = 0
jumlah (tfd * tfd) = 0
jumlah (tfq * tfq) = 14
akar (tfd * tfd) = 0
akar (tfq * tfq) = 3.7416573867739
nilai cosine = jumlah (tfd * tfq) / (akar (tfd * tfd) * akar (tfq * tfq)) =
```

kategori 4:

```
flash ---> tfd=2 ---> tfq=1
animasi ---> tfd=2 ---> tfq=1
desain ---> tfd=1 ---> tfq=1
aplikasi ---> tfd=8 ---> tfq=1
media ---> tfd=2 ---> tfq=1
macromedia ---> tfd=1 ---> tfq=1

jumlah (tfd * tfq) = 16
jumlah (tfd * tfd) = 78
jumlah (tfq * tfq) = 17
akar (tfd * tfd) = 8.8317608663278
akar (tfq * tfq) = 4.1231056256177
nilai cosine = jumlah (tfd * tfq) / (akar (tfd * tfd) * akar (tfq * tfq)) = 0.43938802910487
```

**Gambar4.**Nilai cosinesimilarity kategori 3 dan 4 dokumen termasuk kategori Multimedia

**Gambar5.**Hasil Akhir Pengelompokan

Berdasarkan hasil penghitungan nilai cosine similarity pada sistem diketahui kategori dokumen adalah multimedia. Hal ini menunjukkan bahwa sistem sudah berjalan dengan benar.

Dengan mengambil sampel dokumen sebanyak 5 buah dapat diketahui hasilnya bahwa dengan menggunakan perhitungan nilai cosine similarity secara manual memiliki hasil yang sama dengan perhitungan nilai cosine similarity oleh sistem. Hal yang menunjukkan kesamaan tersebut adalah munculnya kategori yang sama antara perhitungan manual dan perhitungan oleh sistem dalam satu kategori. Hasil yang sama antara perhitungan manual dan perhitungan oleh sistem menunjukkan bahwa penerapan metode cosine similarity pada sistem pengelompokan otomatis telah sesuai dengan keadaan yang sebenarnya.

#### 4.2 PENGUJIAN

Pengujian sistem yang digunakan adalah dengan menggunakan pengujian blackbox. Berikut merupakan tabel pengujian blackbox terhadap sistem:

**Tabel 1.** Pengujian Blackbox

N O	Komponen Uji	Skenario Uji	Hasil yang diharapkan	Hasil Uji
1	Login Admin	Memasukkan username dan password kedalam textbox yang telah disediakan dengan benar dan salah	Sistem menampilkan halaman utama jika username dan password benar. Untuk username dan password yang salah, sistem meminta untuk mengulangi login kembali.	[ v ] Berhasil

NO	Komponen Uji	Skenario Uji	Hasil yang diharapkan	Hasil Uji
2	Input Data Abstraksi dan Pengelompokan Data abstraksi oleh sistem	Mahasiswa dan petugas perpustakaan memasukkan data abstraksi kedalam sistem.	Sistem akan menyimpan data yang dimasukkan mahasiswa dan petugas kemudian secara otomatis sistem mengelompokkan data berdasarkan kategori yang telah ditentukan.	[ v ] Berhasil
3	Manajemen Data Keyword	Petugas perpustakaan memasukkan data keyword, mengubah data keyword dan menghapus data keyword.	Sistem data menyimpan data keyword yang dimasukkan oleh petugas, sistem dapat menyimpan perubahan data yang dilakukan oleh petugas perpustakaan, serta sistem dapat menghapus data yang dipilih oleh petugas perpustakaan.	[ v ] Berhasil
4	Manajemen Data Abstraksi	Petugas perpustakaan mengubah data abstraksi dan menghapus data abstraksi	Sistem dapat menyimpan perubahan yang dilakukan oleh petugas perpustakaan, sistem dapat menghapus data yang dipilih oleh petugas perpustakaan.	[ v ] Berhasil
5	Cetak Bukti Pengumpulan	Mahasiswa mendapatkan bukti pengumpulan setelah memasukkan data dalam sistem.	Sistem akan menampilkan lembar cetak bukti pengumpulan yang dapat disimpan mahasiswa.	[ v ] Berhasil

## V. PENUTUP

### 5.1 KESIMPULAN

Berdasarkan penelitian yang telah dilaksanakan maka diperoleh kesimpulan sebagai berikut:

1. Penerapan metode cosine similarity pada sistem pengelompokan otomatis menghasilkan sebuah sistem yang dapat mengelompokkan dokumen kedalam masing-masing kategoriyaitukategorisisteminformasi, jaringan, *hardware*, dan multimedia.
2. Penentuan kategori dari sebuah dokumen yang telah dilakukan proses perhitungan adalah dengan mengambil nilai cosine similarity yang paling besar terhadap pada sebuah kategori.
3. Berdasarkan hasil sampling yang dilakukan dapat diketahui bahwa faktor

yang mempengaruhi dokumen masuk kedalam sebuah kategori adalah kata-kata penyusun dokumen tersebut.

### 5.2 SARAN

Berdasarkan penelitian dan pembahasan diatas maka penulis menyampaikan saran yaitu:

1. Sistem dapat dikembangkan dengan menambahkan fungsi login terhadap mahasiswa sehingga keamanan sistem dapat meningkat.
2. Sistem dapat ditingkatkan menjadi sistem informasi manajemen pengumpulan laporan kerja praktik, laporan tugas akhir, dan laporan skripsi.

### DAFTAR PUSTAKA

- [1] Purwaingrum, A.K. 2014. “*Pengkategorian Otomatis Dokumen Tugas Akhir D3 Teknik Informatika di Perpustakaan FMIPA UNS*”. Tugas Akhir. Universitas Sebelas Maret. Surakarta.
- [2] Tala, F.Z. 2003. “*A study of Stemming Effects on Information Retrieval in Bahasa Indonesia*”. Master of Logic Project. Universitas van Amsterdam. The Netherlands.
- [3] Purwaingrum, A.K. 2014. “*Pengkategorian Otomatis Dokumen Tugas Akhir D3 Teknik Informatika di Perpustakaan FMIPA UNS*”. Tugas Akhir. Universitas Sebelas Maret. Surakarta.
- [4] Rachman, F H.2011.“*Sistem Temu Kembali Informasi Dalam mesin Pencarian Menggunakan Model Ruang Vektor dan Inverted Index*”. Program Pascasarjana. Universitas Gajah Mada. Yogyakarta.
- [5] Rachman, F H.2011.“*Sistem Temu Kembali Informasi Dalam mesin Pencarian Menggunakan Model Ruang Vektor dan Inverted Index*”. Program Pascasarjana. Universitas Gajah Mada. Yogyakarta.
- [6] Rachman, F H.2011.“*Sistem Temu Kembali Informasi Dalam mesin Pencarian Menggunakan Model Ruang Vektor dan Inverted Index*”. Program Pascasarjana. Universitas Gajah Mada. Yogyakarta.