

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Perpustakaan**

Perpustakaan adalah institusi pengelola karya tulis, karya cetak, atau karya rekam secara profesional dengan sistem yang baku guna memenuhi kebutuhan pendidikan, penelitian, pelestarian, informasi, dan rekreasi para pemustaka (Arsip Nasional Republik Indonesia, 2010:3). Perpustakaan memiliki tugas pokok dalam memenuhi layanan informasi, yaitu:

1. Menghimpun bahan pustaka yang meliputi buku dan nonbuku sebagai sumber informasi.
2. Mengolah dan merawat pustaka.
3. Memberikan layanan bahan pustaka.

Beberapa jenis perpustakaan yang ada saat ini adalah perpustakaan internasional, perpustakaan nasional, perpustakaan umum dan keliling, perpustakaan swasta (pribadi), serta perpustakaan khusus. Jenis koleksi yang umum tersedia di dalam perpustakaan adalah buku, koleksi referensi, jenis serial, brosur, dan materi audio visual. Jenis buku yang dimaksud adalah buku teks, buku penunjang, laporan kerja praktek, laporan tugas akhir, dan tesis.

#### **2.2 *Information Retrieval***

*Information Retrieval* menurut Manning (2009:1) adalah menemukan material (dokumen) dari domain yang tidak terstruktur (bentuk teks) berupa kebutuhan informasi yang memuaskan dari jumlah koleksi yang. Salah satu

contoh *information retrieval* adalah pencarian bahan pustaka oleh mahasiswa di perpustakaan. Perkembangan jaringan *internet* yang sangat pesat semakin membutuhkan *information retrieval* dikarenakan jumlah informasi yang tersedia di dalam jaringan yang semakin membesar. Sehingga dapat disimpulkan bahwa *information retrieval* merupakan pencarian informasi yang berdasarkan pada kata pencarian yang diharapkan mampu memenuhi keinginan pengguna (relevan) dari kumpulan dokumen yang tersedia.

Pada dasarnya *Information Retrieval* berbeda dengan *data retrieval* dimana *data retrieval* menghasilkan data atau dokumen yang tepat sesuai dengan kata kunci yang diberikan. Menurut Herdiyeni (2008) perbedaan *information retrieval* dengan *data retrieval* dapat dilihat pada Tabel 2.1.

Tabel 2.1 Perbedaan *Information Retrieval* dengan *Data Retrieval*

| <i>Information Retrieval</i>   | <i>Data Retrieval</i>   |
|--|---|
| Berhubungan dengan text bahasa umum yang tidak selalu terstruktur dan ada kemungkinan memiliki kerancuan arti. | Berhubungan dengan data, dimana semantik strukturnya sudah terdefiniskan. |
| Semantik sering kali hilang.   | Semantik terdefinisi dengan baik.   |
| Kesalahan kecil masih bisa ditorensi.  | Kesalahan kecil/tunggal dari suatu obyek menunjukkan kegagalan.           |

Menurut Bawono (2012:1) komponen-komponen penting yang berperan di dalam *information retrieval* adalah *indexing*, yang terdiri atas *tokenization*, *stopwords*, dan *stemming*, pembobotan (*weighting*), serta perhitungan tingkat kemiripan dokumen (*similarity measure*). Keseluruhan komponen tersebut tersusun menjadi satu rangkaian dalam proses pencarian dokumen.

### 2.3 Indexing

Menurut Bawono (2012:2) *indexing* adalah proses pembentukan dan pengumpulan kata-kata dalam dokumen elektronik yang akan membentuk sebuah pola pengenalan dokumen tersebut. Dengan membentuk *indexing*, proses *information retrieval* dapat dijalankan dengan cepat karena tersedianya jalur akses ke dalam data atau dokumen.

Untuk melakukan *indexing*, langkah-langkah yang dijalankan adalah:

#### 1. Tokenization

Menurut Manning (2009:19) *tokenization* adalah proses pemecahan suatu kalimat menjadi perkata dimulai dengan membaca dokumen-dokumen yang dimiliki dan memecahnya. Jika tersedia kalimat “Saya belajar *information retrieval*”, maka hasil dari *tokenization* adalah:

|      |         |             |           |
|------|---------|-------------|-----------|
| saya | belajar | information | retrieval |
|------|---------|-------------|-----------|

#### 2. Stopwords

Menurut Bawono (2012:2) proses penghilangan *stopwords* atau kata-kata yang tidak berkaitan langsung dengan kata kunci, seperti kata hubung atau tanda baca dapat mengurangi proses *indexing* dan mengurangi *noise* dalam perhitungan relevansi. Contoh *stopwords* Bahasa Indonesia dapat dilihat pada Tabel 2.2.

Tabel 2.2 Contoh *Stopwords* Bahasa Indonesia

| <b><i>Stopwords</i> Bahasa Indonesia</b> |         |         |        |
|--|---------|---------|--------|
| ada                                      | tentu   | karena  | sedang |
| ini                                      | sudah   | melalui | tetapi |
| adalah                                   | yang    | yaitu   | pula   |
| maupun                                   | kecuali | namun   | juga   |

### 3. *Stemming*

Menurut Agusta (2009) *stemming* adalah proses mentransformasikan kata-kata yang terdapat pada suatu dokumen ke dalam bentuk kata-kata dasarnya (*root word*). Dengan melakukan *stemming*, pencarian kata dengan bentuk yang berbeda tetapi dengan makna yang sama dapat dilakukan dengan cepat. Contohnya kata “mempelajari” dan “pelajaran” memiliki makna yang sama yaitu “ajar”. Dalam melakukan proses *stemming*, metode yang digunakan adalah metode *stemming* Nazief & Adriani.

Untuk membantu dalam memahami *indexing* dan langkah-langkahnya, akan diberikan contoh proses *information retrieval* hingga proses perhitungan relevansi. Contoh proses *indexing* jika tersedia dokumen (D1, D2, dst.) dan kalimat pencarian (*Query/Q*) sebagai berikut:

D.1. Komisi Yudisial dengan universitas resmi menjalin kerjasama dalam memberantas mafia peradilan.

D.2. SAR Trunojoyo mengadakan DIKLAT mereka di Bumi Perkemahan Wisata Air Terjun Mojokerto. Acara itu dibuka langsung oleh pembantu rektor.

- D.3. *Roadshow* Speedy yang di Trunojoyo diisi dengan berbagai acara yaitu seminar internet sehat, *cangkrukan* komunitas, *workshop*, serta lomba band.
- D.4. Pemerintah Kabupaten Pamekasan menghentikan sementara program bantuan beasiswa kepada mahasiswa Pamekasan yang kuliah di Universitas Trunojoyo.
- D.5. Sebanyak 11 orang staf Universitas Trunojoyo melakukan magang di Fakultas Teknik Industri UII.
- D.6. Perpustakaan Universitas Airlangga pada minggu ini kedatangan tamu dari 2 perguruan tinggi, yaitu staf Perpustakaan Trunojoyo dan staf Perpustakaan Gunadharma.

Q. Universitas Trunojoyo.

Pada contoh dokumen yang tersedia dan kalimat pencarian, terdapat total dokumen yang terlibat sebanyak 7 buah ( $N=7$ ) dengan rincian 6 dokumen yang tersedia dan 1 kalimat pencarian. Hasil *tokenization*, *stopwords*, dan *stemming* dari dokumen yang disediakan dengan kalimat pencarian dapat dilihat pada Tabel 2.3.

Tabel 2.3 Hasil *Tokenization*, *Stopwords*, dan *Stemming*.

| Dokumen | Term Yang Mewakili Dokumen  |
|---------|---|
| Q       | universitas trunojoyo   |
| D1      | komisi yudisial universitas jalin kerjasama berantas mafia adil                         |
| D2      | sar trunojoyo diklat bumi kemah wisata air terjun mojokerto bantu rektor                |
| D3      | roadshow speedy trunojoyo seminar internet sehat cangkruk komunitas workshop lomba band |

| <b>Dokumen</b> | <b>Term Yang Mewakili Dokumen</b>   |
|----------------|---|
| D4             | perintah kabupaten pamekasan henti program bantu beasiswa mahasiswa pamekasan universitas trunojoyo                           |
| D5             | 11 staf universitas trunojoyo magang fakultas teknik industri uii   |
| D6             | pustaka universitas airlangga datang tamu 2 guru tinggi staf pustaka universitas trunojoyo staf pustaka universitas gunadarma |

Dari Tabel 2.3 dapat dilihat bahwa masing-masing dokumen memiliki kata-kata kunci dan kata-kata yang berkaitan dengan kata kunci itu sendiri dan sudah menjadi kata dasar (*root word*). Langkah selanjutnya adalah membentuk suatu *indexing* yang dapat dilihat pada Tabel 2.4.

Tabel 2.4 Hasil *Indexing*

| <b>Term</b> | <b>Dokumen ID</b> | <b>Term</b> | <b>Dokumen ID</b> | <b>Term</b> | <b>Dokumen ID</b> |
|-------------|-------------------|-------------|-------------------|-------------|-------------------|
| Universitas | Q, 1, 4, 5, 6     | Bantu       | 2                 | Bantu       | 4                 |
| Trunojoyo   | Q, 2, 3, 4, 5, 6  | Rektor      | 2                 | Mahasiswa   | 4                 |
| Komisi      | 1                 | Roadshow    | 3                 | 11          | 5                 |
| Yudisial    | 1                 | Speedy      | 3                 | Staf        | 5                 |
| Jalin       | 1                 | Seminar     | 3                 | Magang      | 5                 |
| Kerjasama   | 1                 | Internet    | 3                 | Fakultas    | 5                 |
| Berantas    | 1                 | Sehat       | 3                 | Teknik      | 5                 |
| Mafia       | 1                 | Cangkruk    | 3                 | Industri    | 5                 |
| Adil        | 1                 | Komunitas   | 3                 | Uii         | 5                 |
| Sar         | 2                 | Workshop    | 3                 | Pustaka     | 6                 |
| Diklat      | 2                 | Lomba       | 3                 | Tamu        | 6                 |
| Bumi        | 2                 | Band        | 3                 | 2           | 6                 |
| Kemah       | 2                 | Perintah    | 4                 | Guru        | 6                 |
| Wisata      | 2                 | Kabupaten   | 4                 | Tinggi      | 6                 |
| Air         | 2                 | Pamekasan   | 4                 | Gunadarma   | 6                 |
| Terjun      | 2                 | Henti       | 4                 |             |                   |
| Mojokerto   | 2                 | Program     | 4                 |             |                   |

## 2.4 Pembobotan (*Term Weighting*)

Menurut Manning (2009:118), pembobotan (*term weighting*) adalah pemberian nilai terhadap istilah (*term*) yang terdapat dalam dokumen. Semakin sering suatu istilah muncul dalam suatu dokumen, semakin tinggi bobot dokumen tersebut (*Term Frequency/TF*). Metode pembobotan yang digunakan merupakan kombinasi dari *Term Frequency* (TF) dan *Inverted Document Frequency* (IDF). *Inverted Term Frequency* merupakan konsep kebalikan dari *Term Frequency* dimana semakin banyak dokumen yang mengandung sebuah istilah, maka semakin kecil bobot istilah tersebut. Rumus pembobotan TF-IDF menurut Manning dapat dilihat pada Persamaan 2.1, 2.2, dan 2.3.

$$TF_{d,t} = f_{d,t} \dots\dots\dots(\text{Persamaan 2.1})$$

$$IDF_t = \log(N/df_t) \dots\dots\dots(\text{Persamaan 2.2})$$

$$W_{d,t} = TF_{d,t} * IDF_t \dots\dots\dots(\text{Persamaan 2.3})$$

Dimana:

- d = dokumen ke-d
- t = *term* (istilah) ke-t dari dokumen
- $f_{d,t}$  = jumlah kemunculan *term* atau istilah t pada dokumen d
- N = jumlah seluruh dokumen di dalam *database*
- $df_t$  = jumlah dokumen yang memiliki istilah t.

Untuk contoh pembobotan, dapat melanjutkan hasil *indexing* yang terdapat pada Tabel 2.4. Pada Tabel 2.5 disusun perhitungan frekuensi

kemunculan dari masing-masing *term* yang terdapat pada dokumen yang tersedia dan kemudian dilanjutkan dengan pembobotan.

Tabel 2.5 Perhitungan Frekuensi Kemunculan *Term* dan Pembobotan

| <i>Term</i> | TF |    |    |    |    |    |    |    | IDF   | W = TF * IDF |       |       |       |       |       |       |
|-------------|----|----|----|----|----|----|----|----|-------|--------------|-------|-------|-------|-------|-------|-------|
|             | Q  | D1 | D2 | D3 | D4 | D5 | D6 | df |       | Q            | D1    | D2    | D3    | D4    | D5    | D6    |
| Universitas | 1  | 1  |    |    | 1  | 1  | 1  | 5  | 0,146 | 0,146        | 0,146 |       |       | 0,146 | 0,146 | 0,146 |
| Trunojoyo   | 1  |    | 1  | 1  | 1  | 1  | 1  | 6  | 0,066 | 0,066        |       | 0,066 | 0,066 | 0,066 | 0,066 | 0,066 |
| Komisi      |    | 1  |    |    |    |    |    | 1  | 0,845 |              | 0,845 |       |       |       |       |       |
| Yudisial    |    | 1  |    |    |    |    |    | 1  | 0,845 |              | 0,845 |       |       |       |       |       |
| Jalin       |    | 1  |    |    |    |    |    | 1  | 0,845 |              | 0,845 |       |       |       |       |       |
| Kerjasama   |    | 1  |    |    |    |    |    | 1  | 0,845 |              | 0,845 |       |       |       |       |       |
| Berantas    |    | 1  |    |    |    |    |    | 1  | 0,845 |              | 0,845 |       |       |       |       |       |
| Mafia       |    | 1  |    |    |    |    |    | 1  | 0,845 |              | 0,845 |       |       |       |       |       |
| Adil        |    | 1  |    |    |    |    |    | 1  | 0,845 |              | 0,845 |       |       |       |       |       |
| Sar         |    |    | 1  |    |    |    |    | 1  | 0,845 |              |       | 0,845 |       |       |       |       |
| Diklat      |    |    | 1  |    |    |    |    | 1  | 0,845 |              |       | 0,845 |       |       |       |       |
| Bumi        |    |    | 1  |    |    |    |    | 1  | 0,845 |              |       | 0,845 |       |       |       |       |
| Kemah       |    |    | 1  |    |    |    |    | 1  | 0,845 |              |       | 0,845 |       |       |       |       |
| Wisata      |    |    | 1  |    |    |    |    | 1  | 0,845 |              |       | 0,845 |       |       |       |       |
| Air         |    |    | 1  |    |    |    |    | 1  | 0,845 |              |       | 0,845 |       |       |       |       |
| Terjun      |    |    | 1  |    |    |    |    | 1  | 0,845 |              |       | 0,845 |       |       |       |       |
| Mojokerto   |    |    | 1  |    |    |    |    | 1  | 0,845 |              |       | 0,845 |       |       |       |       |
| Bantu       |    |    | 1  |    |    |    |    | 1  | 0,845 |              |       | 0,845 |       |       |       |       |
| Rektor      |    |    | 1  |    |    |    |    | 1  | 0,845 |              |       | 0,845 |       |       |       |       |
| Roadshow    |    |    |    | 1  |    |    |    | 1  | 0,845 |              |       |       | 0,845 |       |       |       |
| Speedy      |    |    |    | 1  |    |    |    | 1  | 0,845 |              |       |       | 0,845 |       |       |       |
| Seminar     |    |    |    | 1  |    |    |    | 1  | 0,845 |              |       |       | 0,845 |       |       |       |
| Internet    |    |    |    | 1  |    |    |    | 1  | 0,845 |              |       |       | 0,845 |       |       |       |
| Sehat       |    |    |    | 1  |    |    |    | 1  | 0,845 |              |       |       | 0,845 |       |       |       |
| Cangkruk    |    |    |    | 1  |    |    |    | 1  | 0,845 |              |       |       | 0,845 |       |       |       |
| Komunitas   |    |    |    | 1  |    |    |    | 1  | 0,845 |              |       |       | 0,845 |       |       |       |
| Workshop    |    |    |    | 1  |    |    |    | 1  | 0,845 |              |       |       | 0,845 |       |       |       |
| Lomba       |    |    |    | 1  |    |    |    | 1  | 0,845 |              |       |       | 0,845 |       |       |       |
| Band        |    |    |    | 1  |    |    |    | 1  | 0,845 |              |       |       | 0,845 |       |       |       |
| Perintah    |    |    |    |    | 1  |    |    | 1  | 0,845 |              |       |       |       | 0,845 |       |       |
| Kabupaten   |    |    |    |    | 1  |    |    | 1  | 0,845 |              |       |       |       | 0,845 |       |       |
| Pamekasan   |    |    |    |    | 2  |    |    | 1  | 0,845 |              |       |       |       | 1,69  |       |       |



| Term      | TF |    |    |    |    |    |    |    | IDF   | W = TF * IDF |    |    |    |       |       |       |
|-----------|----|----|----|----|----|----|----|----|-------|--------------|----|----|----|-------|-------|-------|
|           | Q  | D1 | D2 | D3 | D4 | D5 | D6 | df |       | Q            | D1 | D2 | D3 | D4    | D5    | D6    |
| Henti     |    |    |    |    | 1  |    |    | 1  | 0,845 |              |    |    |    | 0,845 |       |       |
| Program   |    |    |    |    | 1  |    |    | 1  | 0,845 |              |    |    |    | 0,845 |       |       |
| Bantu     |    |    |    |    | 1  |    |    | 1  | 0,845 |              |    |    |    | 0,845 |       |       |
| Mahasiswa |    |    |    |    | 1  |    |    | 1  | 0,845 |              |    |    |    | 0,845 |       |       |
| 11        |    |    |    |    |    | 1  |    | 1  | 0,845 |              |    |    |    |       | 0,845 |       |
| Staf      |    |    |    |    |    | 1  | 2  | 2  | 0,544 |              |    |    |    |       | 0,544 | 1,088 |
| Magang    |    |    |    |    |    | 1  |    | 1  | 0,845 |              |    |    |    |       | 0,845 |       |
| Fakultas  |    |    |    |    |    | 1  |    | 1  | 0,845 |              |    |    |    |       | 0,845 |       |
| Teknik    |    |    |    |    |    | 1  |    | 1  | 0,845 |              |    |    |    |       | 0,845 |       |
| Industri  |    |    |    |    |    | 1  |    | 1  | 0,845 |              |    |    |    |       | 0,845 |       |
| Uii       |    |    |    |    |    | 1  |    | 1  | 0,845 |              |    |    |    |       | 0,845 |       |
| Pustaka   |    |    |    |    |    |    | 1  | 1  | 0,845 |              |    |    |    |       |       | 0,845 |
| Tamu      |    |    |    |    |    |    | 3  | 1  | 0,845 |              |    |    |    |       |       | 2,535 |
| 2         |    |    |    |    |    |    | 1  | 1  | 0,845 |              |    |    |    |       |       | 0,845 |
| Guru      |    |    |    |    |    |    | 1  | 1  | 0,845 |              |    |    |    |       |       | 0,845 |
| Tinggi    |    |    |    |    |    |    | 1  | 1  | 0,845 |              |    |    |    |       |       | 0,845 |
| Gunadarma |    |    |    |    |    |    | 1  | 1  | 0,845 |              |    |    |    |       |       | 0,845 |

## 2.5 Similarity Measure

Metode yang digunakan untuk mengukur kemiripan (*similarity measure*) menggunakan metode *Cosine Similarity*. Menurut Bawono (2012:3), metode *cosine similarity* menerapkan konsep trigonometri kosinus yaitu mengukur kemiripan antara dua objek berdasarkan sudut yang dibentuk oleh kedua objek tersebut. Semakin besar nilai kosinus (mendekati 1) maka semakin mirip antara *query* pencarian dengan dokumen yang ada pada *database*. Semakin kecil nilai kosinus (mendekati 0) maka semakin tidak mirip antara pencarian dengan dokumen yang ada. Apabila nilai kosinus adalah 1, maka kemiripan dokumen 100%. Sedangkan nilai 0, maka kemiripan dokumen 0%. Rumus perhitungan kemiripan kosinus atau *cosine similarity* dapat dilihat pada Persamaan 2.4.

$$\text{similarity}(\vec{d}_j, \vec{q}) = \frac{\vec{d}_j \cdot \vec{q}}{|\vec{d}_j| \cdot |\vec{q}|} = \frac{\sum_{i=1}^t (w_{ij} \cdot w_{iq})}{\sqrt{\sum_{i=1}^t (w_{ij})^2 \cdot \sum_{i=1}^t (w_{iq})^2}} \dots\dots\dots(2.4)$$

Dimana:

- d = dokumen
- t = *term* (istilah)
- q = *query*
- $w_{ij}$  = TF-IDF kata ke-i dari dokumen ke-j
- $w_{iq}$  = TF-IDF kata ke-i dari *query*

Untuk mengetahui contoh proses perhitungan kemiripan dengan menggunakan metode *cosine similarity*, dapat dilihat pada Tabel 2.6 dimana proses ini melanjutkan hasil perhitungan pembobotan yang terdapat pada Tabel 2.5 sebelumnya.

Tabel 2.6 Perhitungan *Cosine Similarity*

| <i>Term</i> | Wij * Wiq |       |       |       |       |       | Panjang Vektor |       |       |       |       |       |       |
|-------------|-----------|-------|-------|-------|-------|-------|----------------|-------|-------|-------|-------|-------|-------|
|             | D1        | D2    | D3    | D4    | D5    | D6    | Q              | D1    | D2    | D3    | D4    | D5    | D6    |
| Universitas | 0,021     |       |       | 0,021 | 0,021 | 0,021 | 0,021          | 0,021 |       |       | 0,021 | 0,021 | 0,021 |
| Trunojoyo   |           | 0,004 | 0,004 | 0,004 | 0,004 | 0,004 | 0,004          | 0     | 0,004 | 0,004 | 0,004 | 0,004 | 0,004 |
| Komisi      |           |       |       |       |       |       |                | 0,714 |       |       |       |       |       |
| Yudisial    |           |       |       |       |       |       |                | 0,714 |       |       |       |       |       |
| Jalin       |           |       |       |       |       |       |                | 0,714 |       |       |       |       |       |
| Kerjasama   |           |       |       |       |       |       |                | 0,714 |       |       |       |       |       |
| Berantas    |           |       |       |       |       |       |                | 0,714 |       |       |       |       |       |
| Mafia       |           |       |       |       |       |       |                | 0,714 |       |       |       |       |       |
| Adil        |           |       |       |       |       |       |                | 0,714 |       |       |       |       |       |
| Sar         |           |       |       |       |       |       |                |       | 0,714 |       |       |       |       |
| Diklat      |           |       |       |       |       |       |                |       | 0,714 |       |       |       |       |
| Bumi        |           |       |       |       |       |       |                |       | 0,714 |       |       |       |       |
| Kemah       |           |       |       |       |       |       |                |       | 0,714 |       |       |       |       |

| <i>Term</i>     | Wij * Wiq    |              |              |              |              |              | Panjang Vektor |              |              |              |              |              |               |
|-----------------|--------------|--------------|--------------|--------------|--------------|--------------|----------------|--------------|--------------|--------------|--------------|--------------|---------------|
|                 | D1           | D2           | D3           | D4           | D5           | D6           | Q              | D1           | D2           | D3           | D4           | D5           | D6            |
| Wisata          |              |              |              |              |              |              |                |              | 0,714        |              |              |              |               |
| Air             |              |              |              |              |              |              |                |              | 0,714        |              |              |              |               |
| Terjun          |              |              |              |              |              |              |                |              | 0,714        |              |              |              |               |
| Mojokerto       |              |              |              |              |              |              |                |              | 0,714        |              |              |              |               |
| Bantu           |              |              |              |              |              |              |                |              | 0,714        |              |              |              |               |
| Rektor          |              |              |              |              |              |              |                |              | 0,714        |              |              |              |               |
| Roadshow        |              |              |              |              |              |              |                |              |              | 0,714        |              |              |               |
| Speedy          |              |              |              |              |              |              |                |              |              | 0,714        |              |              |               |
| Seminar         |              |              |              |              |              |              |                |              |              | 0,714        |              |              |               |
| Internet        |              |              |              |              |              |              |                |              |              | 0,714        |              |              |               |
| Sehat           |              |              |              |              |              |              |                |              |              | 0,714        |              |              |               |
| Cangkruk        |              |              |              |              |              |              |                |              |              | 0,714        |              |              |               |
| Komunitas       |              |              |              |              |              |              |                |              |              | 0,714        |              |              |               |
| Workshop        |              |              |              |              |              |              |                |              |              | 0,714        |              |              |               |
| Lomba           |              |              |              |              |              |              |                |              |              | 0,714        |              |              |               |
| Band            |              |              |              |              |              |              |                |              |              | 0,714        |              |              |               |
| Perintah        |              |              |              |              |              |              |                |              |              |              | 0,714        |              |               |
| Kabupaten       |              |              |              |              |              |              |                |              |              |              | 0,714        |              |               |
| Pamekasan       |              |              |              |              |              |              |                |              |              |              | 2,856        |              |               |
| Henti           |              |              |              |              |              |              |                |              |              |              | 0,714        |              |               |
| Program         |              |              |              |              |              |              |                |              |              |              | 0,714        |              |               |
| Bantu           |              |              |              |              |              |              |                |              |              |              | 0,714        |              |               |
| Mahasiswa       |              |              |              |              |              |              |                |              |              |              | 0,714        |              |               |
| 11              |              |              |              |              |              |              |                |              |              |              |              | 0,714        |               |
| Staf            |              |              |              |              |              |              |                |              |              |              |              | 0,544        | 1,183         |
| Magang          |              |              |              |              |              |              |                |              |              |              |              | 0,714        |               |
| Fakultas        |              |              |              |              |              |              |                |              |              |              |              | 0,714        |               |
| Teknik          |              |              |              |              |              |              |                |              |              |              |              | 0,714        |               |
| Industri        |              |              |              |              |              |              |                |              |              |              |              | 0,714        |               |
| Uii             |              |              |              |              |              |              |                |              |              |              |              | 0,714        |               |
| Pustaka         |              |              |              |              |              |              |                |              |              |              |              |              | 0,714         |
| Tamu            |              |              |              |              |              |              |                |              |              |              |              |              |               |
| 2               |              |              |              |              |              |              |                |              |              |              |              |              | 0,714         |
| Guru            |              |              |              |              |              |              |                |              |              |              |              |              | 0,714         |
| Tinggi          |              |              |              |              |              |              |                |              |              |              |              |              | 0,714         |
| Gunadarma       |              |              |              |              |              |              |                |              |              |              |              |              | 0,714         |
| $\Sigma$        | <b>0,021</b> | <b>0,004</b> | <b>0,004</b> | <b>0,026</b> | <b>0,026</b> | <b>0,026</b> | <b>0,025</b>   | <b>5,019</b> | <b>7,144</b> | <b>7,144</b> | <b>7,165</b> | <b>4,853</b> | <b>11,204</b> |
| $\sqrt{\Sigma}$ |              |              |              |              |              |              | <b>0,16</b>    | <b>2,240</b> | <b>2,672</b> | <b>2,672</b> | <b>2,677</b> | <b>2,203</b> | <b>3,347</b>  |

Setelah komponen pembilang dan penyebut pada rumus *cosine similarity* didapatkan dari Tabel 2.6, langkah berikutnya adalah menghitung nilai kosinus yang kemudian menjadi bentuk persentase. Hasil perhitungan kosinus antara *query* pencarian dengan dokumen-dokumen yang tersedia adalah sebagai berikut:

1.  $\text{Cos (Q, D1)} = 0,021/(0,16*2,240) = 0,059$ , sehingga kemiripannya adalah 5,9%.
2.  $\text{Cos (Q, D2)} = 0,004/(0,16*2,672) = 0,010$ , sehingga kemiripannya adalah 1%.
3.  $\text{Cos (Q, D3)} = 0,004/(0,16*2,672) = 0,010$ , sehingga kemiripannya adalah 1%.
4.  $\text{Cos (Q, D4)} = 0,026/(0,16*2,677) = 0,059$ , sehingga kemiripannya adalah 5,9%.
5.  $\text{Cos (Q, D5)} = 0,026/(0,16*2,203) = 0,073$ , sehingga kemiripannya adalah 7,3%.
6.  $\text{Cos (Q, D6)} = 0,026/(0,16*3,347) = 0,048$ , sehingga kemiripannya adalah 4,8%.

Setelah menghitung kesamaan kosinus, maka langkah selanjutnya adalah mengurutkan dokumen dari yang memiliki tingkat kemiripan tertinggi ke yang terendah. Berdasarkan hasil perhitungan, urutan dokumen yang dihasilkan atau ditampilkan adalah D5, D4, D1, D6, D2, dan D3.

## 2.6 Teknik Pengambilan Sampel (*Sampling*)

Dalam penelitian yang melibatkan jumlah populasi yang cukup besar, diperlukan sebagian dari populasi tersebut yang akan diteliti sebagai sampel.

Menurut Hendriyadi (2010) sampel adalah perwakilan dari populasi sehingga hasil penelitian yang diperoleh dari sampel dapat digeneralisasikan pada populasi. Ukuran sampel yang umum digunakan lebih dari 30 dan kurang dari 500. Namun jumlah sampel yang dibutuhkan bergantung pada tingkat kesalahan yang diinginkan peneliti. Batas maksimal kesalahan yang digunakan 5% atau tingkat ketelitian 95%.

Teknik yang umum digunakan untuk menentukan jumlah sampel adalah dengan menggunakan rumus Slovin. Rumus Slovin digunakan untuk menentukan jumlah minimal sampel ( $n$ ) dari populasi ( $N$ ) yang ada dengan tingkat kesalahan ( $e$ ) yang dikehendaki. Adapun rumus Slovin dapat dilihat pada Persamaan 2.5.

Dimana:

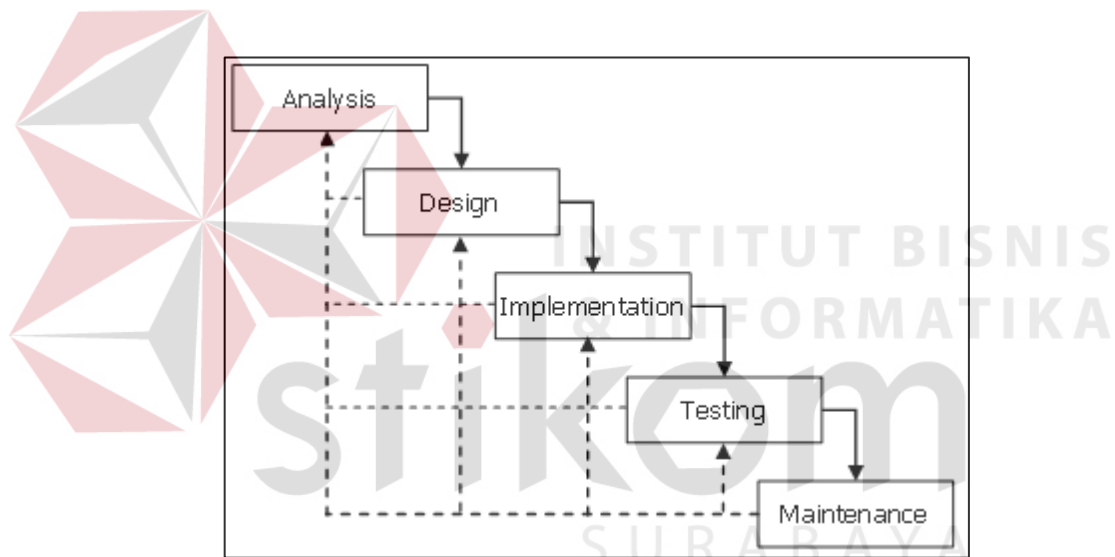
- $n$  = nilai sampel minimal.
- $N$  = jumlah populasi.
- $e$  = batas nilai toleransi kesalahan.

$$n = \frac{N}{1 + Ne^2} \dots \dots \dots (2.5)$$

## 2.7 System Development Life Cycle (SDLC)

Daur hidup pengembangan sistem atau SDLC merupakan siklus dan tahapan-tahapan yang terjadi dalam pengembangan sistem. Model SDLC yang umum digunakan hingga kini adalah metode *Waterfall*. Menurut Bassil (2012), model *Waterfall* adalah proses pengembangan perangkat lunak (*software*) yang berurutan di mana kemajuan proses dianggap mengalir ke bawah seperti air terjun

yang melalui tahapan-tahapan yang harus ditempuh untuk keberhasilan dalam pengembangan *software*. Bassil (2012) juga memaparkan bahwa model *Waterfall* ini pada awalnya dikemukakan oleh Winston W. Royce pada tahun 1970 sebagai penggambaran praktis dalam pengembangan *software*. Model *Waterfall* menjelaskan urutan-urutan tahapan yang harus diselesaikan, dan perpindahan menuju tahap berikutnya hanya ketika tahapan berikutnya benar-benar dijalankan. Oleh karena itu model *Waterfall* bersifat rekursif yaitu setiap tahapan dapat dapat diulang tanpa henti hingga benar-benar sempurna.



Gambar 2.1 SDLC Model *Waterfall*

Pada Gambar 2.1 menunjukkan tahapan-tahapan dalam SDLC dengan menerapkan model *Waterfall*. Pada intinya *Waterfall* terdiri atas tahapan-tahapan sebagai berikut:

1. Analisis

Dalam tahapan analisis terdapat pembahasan mengenai tugas-tugas yang akan dikerjakan oleh pengguna *software*, informasi apa saja yang akan

diberikan ke pengguna, dan proses-proses penting apa saja yang harus dibuat.

## 2. Perancangan

Pada tahapan perancangan terdapat pembahasan mengenai struktur dan bentuk dari *software* yang akan dibangun, seperti struktur basis data, aliran data, alur kerja *software*, dan tampilan tatap muka *software*.

## 3. Implementasi

Tahapan implementasi merupakan menerjemahkan konsep-konsep yang telah dirancang sebelumnya menjadi *software* dengan menerapkan dasar-dasar ilmu pemrograman.

## 4. Pengujian

Pengujian dilaksanakan untuk memastikan bahwa *software* yang sedang dibangun sudah sesuai dengan perencanaan awal, tidak ada kesalahan, dan terintegrasi dengan baik.

## 5. Perawatan

Perawatan merupakan bentuk pemeliharaan terhadap *software* yang telah dibangun dengan cara memberikan panduan kepada pengguna mengenai penggunaan *software* atau dengan cara memberikan bantuan langsung kepada pengguna dari pengembang *software*.