



## Hak Cipta Dilindungi Undang-Undang

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

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

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

## BAB II

### LANDASAN TEORI

#### 2.1 Al-Quran

Menurut Kamus Besar Bahasa Indonesia (KBBI) Al-Quran adalah kitab suci umat Islam yang berisi firman Allah yang diturunkan kepada Nabi Muhammad saw. dengan perantaraan malaikat Jibril untuk dibaca, dipahami, dan diamalkan sebagai petunjuk atau pedoman hidup bagi umat manusia sedangkan menurut bahasa, “Qur’an” berarti “bacaan”, pengertian seperti ini dikemukakan dalam Al-Qur’an sendiri yakni dalam surat Al-Qiyamah, ayat 17-18 “*Sesungguhnya mengumpulkan Al-Qur’an (di dalam dadamu) dan (menetapkan) bacaannya (pada lidahmu) itu adalah tanggungan kami. (Karena itu), jika kami telah membacakannya, hendaklah kamu ikuti bacaannya*”. Adapun menurut istilah Al-Qur’an berarti: “*Kalam Allah yang merupakan mu’jizat yang diturunkan kepada nabi Muhammad, yang disampaikan secara mutawatir dan membacanya adalah ibadah*”

fungsi Al-Qur’an yang sebagaimana tersurat nama-namanya dalam Al-Qur’an adalah sebagai berikut (Latif, 2014)

1. Al-Huda (petunjuk), Dalam Al-Qur'an terdapat tiga kategori tentang posisi Al-Qur'an sebagai petunjuk. Pertama, petunjuk bagi manusia secara umum. Kedua, Al-Qur'an adalah petunjuk bagi orang-orang bertakwa. Ketiga, petunjuk bagi orang-orang yang beriman.
2. Al-Furqon (pemisah), Dalam Al-Qur'an dikatakan bahwa ia adalah ugeran untuk membedakan dan bahkan memisahkan antara yang hak dan yang batil, atau antara yang benar dan yang salah.
3. Al-Asyifa (obat). Dalam Al-Qur'an dikatakan bahwa ia berfungsi sebagai obat bagi penyakit-penyakit yang ada dalam dada (mungkin yang dimaksud di sini adalah penyakit Psikologis)



4. Al-Mau'izah (nasihat), Di dalam Al-Qur'an di katakan bahwa ia berfungsi sebagai penasihat bagi orang-orang yang bertakwa

## 2.2 Aplikasi Al-Quran Indonesia

Aplikasi Al-Quran Indonesia adalah suatu aplikasi mobile yang menyediakan bacaan Al-quran lengkap dengan terjemahan indonesia. Pilihan terjemahan sudah berdasarkan kemenAg-RI dan tafsir Al-jalalain indonesia (<https://play.google.com/store/apps/details?id=com.andi.alquran.id>).

## 2.3 Temu Kembali Informasi

Temu kembali informasi (*Information Retrieval*) digunakan untuk menemukan kembali informasi yang terkumpul dari dokumen-dokumen yang memudahkan pengguna mendapatkan informasi yang relevan. sistem temu kembali informasi berbeda dengan sistem temu kembali data. Tujuan utama dari sistem temu kembali data adalah menentukan dokumen kata kunci yang sesuai dengan query yang dicari dari sekumpulan data. Tujuan utama sistem temu kembali informasi adalah menemukan kembali informasi yang relevan sesuai dengan query di cari dan meminimalisir menemukan dokumen yang tidak relevan.

Menurut Lancaster dikutip oleh (Farisi, 2015) fungsi utama sistem temu kembali informasi dapat disimpulkan sebagai berikut:

1. Untuk menganalisis isi sumber informasi suatu dokumen.
2. Mempresentasikan isi sumber informasi dengan cara tertentu yang memungkinkan untuk ditemukan dengan pernyataan (*query* pengguna).
3. Mempresentasikan pernyataan (*query*) pengguna dengan cara tertentu yang memungkinkan untuk dipertemukan dengan sumber informasi yang terdapat dalam basis data perpustakaan.
4. Mempertemukan pernyataan pencarian dengan data yang tersimpan dalam basis data.
5. Menemu-balikkan informasi yang relevan.



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

6. Menyempurnakan untuk kerja sistem berdasarkan umpan balik yang diberikan oleh pengguna.

Ada beberapa istilah penting yang berkaitan dengan Information Retrieval diantaranya adalah: (Wulandari, 2013).

1. Korpus (*corpus*)

Korpus adalah kumpulan teks dalam kumpulan dari potongan-potongan teks bahasa dalam bentuk elektronik, dipilih sesuai dengan kriteria eksternal untuk mewakili suatu ragam bahasa sebagai sumber data penelitian linguistik.

2. Kueri (*query*)

Kueri adalah inputan informasi yang diinginkan pengguna. Kueri bisa berupa kata kunci yang diberikan terhadap sistem untuk memperoleh hasil sesuai dengan yang diinginkan pengguna. Defenisi kueri secara formal adalah kumpulan spesifikasi yang digunakan untuk menggali data yang ada pada basis data.

3. Relevansi

Relevansi adalah tingkat kesesuaian dari sebuah dokumen dengan kueripengguna. Relevansi dihitung berdasarkan metode yang digunakan pada penelitian sistem temu kembali informasi.

4. Rangkaian (*peringkat*)

Rangkaian adalah peringkat yang diperoleh oleh berbagai dokumen yang mengacu pada kueri pengguna.

5. Term (kata)

kata adalah kata yang memiliki arti yang terdapat pada dokumen dan kueri. Dari semua kata-kata yang merupakan kata umum akan dihilangkan sehingga yang tersisa benar-benar kata yang berhubungan dengan isi dokumen.

## 2.4 Arsitektur Sistem Temu Kembali

Secara umum ada 2 tahapan dalam sistem temu kembali informasi yaitu tahapan *preprocessing* dokumen dan *query* yang mana hasil dari tahapan ini akan

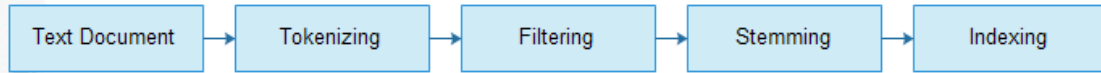
## Hak Cipta Dilindungi Undang-Undang

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.

menghasilkan *inverted index* dapat dilihat pada Gambar 2.1. Dan tahap kedua adalah penerapan metode tertentu untuk menghitung kedekatan dokumen dengan *query* yang dimasukkan oleh pengguna.



**Gambar 2.1 Tahapan *Indexing***

### 2.4.1 *Indexing*

Menurut (Hasuli, 2015), *indexing* dokumen merupakan salah satu tugas pokok dari sistem temu kembali. Karena hasil dari *indexing* tersebut adalah himpunan kata yang menunjukkan isi atau pokok dari dokumen tahapan dari *Indexing* (Gambar 2.1).

#### 1. Teks dokumen

kumpulan korpus dari potongan-potongan teks bahasa dalam bentuk elektronik, dipilih sesuai dengan kriteria eksternal untuk mewakili suatu ragam bahasa sebagai sumber data penelitian linguistik.

#### 2. Tokenisasi

Tokenisasi adalah tahapan untuk menjadikan seluruh kata yang ada di dalam dokumen menjadi token atau potongan kata tunggal. Pada tahap ini juga dilakukan penghapusan tanda baca dan mengubah semua kata menjadi huruf kecil.

#### 3. *Filtering*

Pada tahap *filtering* kita menyaring kata-kata yang telah melalui tahap tokenisasi sehingga kita dapat mengambil kata-kata yang identik dengan dokumen itu. Pada tahapan ini untuk menentukan daftar kata apa saja yang di filter dapat menggunakan daftar *stoplist* atau *wordlist*. *Stoplist* adalah kata umum tidak layak untuk dijadikan kata kunci dalam pencarian dokumen sedangkan *wordlist* adalah kata yang mungkin di jadikan kata kunci dalam pencarian dokumen.



**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. *Stemming*

*Stemming* adalah proses pembuangan imbuhan dalam kata sehingga menjadi kata dasar. Proses ini penting dalam sistem temu kembali karena salah satu dari tahapan untuk meningkatkan kinerja dalam sistem temu kembali tersebut. Pada penelitian ini, menggunakan algoritma *Enhanced confix stripping* yang memiliki presentase keakuratan (presisi) yang tinggi.

##### 2.4.2 *Query*

Dalam suatu pencarian di sistem temu kembali pengguna akan menggunakan *query*. *Query* adalah inputan informasi yang diinginkan pengguna. *query* bisa berupa kata kunci yang diberikan terhadap sistem untuk memperoleh hasil sesuai dengan yang diinginkan pengguna. Defenisi *query* secara formal adalah kumpulan spesifikasi yang digunakan untuk menggali data yang ada pada basis data. (Wulandari, 2013)

##### 2.4.3 **Memeriksa Sinonim**

Setelah kata melewati tahap stemming, kata sinonim tersebut akan diperiksa sesuai dengan kamus tesaurus bahasa Indonesia. Tahapan ini di lakukan pada saat pengguna memasukkan *query*.

##### 2.4.4 **Tesaurus**

Menurut buku Tesaurus (Pusat Bahasa Departemen Pendidikan nasional, 2008), Tesaurus berisi seperangkat kata yang saling bertalian maknanya. Tesaurus adalah buku sinonim yang termasuk karya terkait yang memiliki hal yang hampir sama. Tesaurus dibedakan dari kamus. Di dalam kamus dapat dicari informasi tentang makna kata, sedangkan di dalam tesaurus dapat dicari kata yang akan digunakan untuk mengungkapkan gagasan pengguna. Dengan demikian, penggunaan tesaurus dapat membantu dalam mengungkapkan atau mengekspresikan gagasan sesuai dengan apa yang dimaksud. Misalnya, pencarian kata lain untuk kata hewan, pengguna tesaurus dapat mencarinya pada kata hewan. Contohnya sebagai berikut: hewan binatang, dabat, fauna, sato, satwa



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.

### 2.4.5 Pembobotan Dokumen

Pembobotan kata di gunakan untuk menghitung kemiripan *query* yang di gunakan pengguna dengan koleksi dokumen dengan rumus sebagai sebagai berikut:

$$Pembobotan = TF \times IDF \dots\dots\dots(2.1)$$

$$IDF = \log \frac{n}{df} \dots\dots\dots(2.2)$$

Dimana :

- n = Jumlah dokumen
- df = Banyaknya dokumen dimana suatu kata muncul
- TF = frekuensi kemunculan kata pada dokumen
- IDF = hubungan ketersediaan sebuah kata dalam seluruh dokumen

Penjelasan skema dalam pembobotan kata diantaranya:

#### 2.4.5.1 Term Frequency (TF)

Empat cara yang dapat digunakan untuk memperoleh nilai *term frequency* (tf), yaitu (Garcia, 2008 yang dikutip Sucipto, 2014):

- a. *Raw term frequency.*  
 Nilai tf sebuah *term* diperoleh berdasarkan jumlah kemunculan *term* tersebut dalam dokumen. Contohnya, jika suatu *term* muncul sebanyak tiga kali dalam suatu dokumen maka, nilai *tf term* tersebut adalah 3.
- b. *Logarithm term frequency.*  
 Hal ini untuk menghindari dominasi dokumen yang mengandung sedikit *term* dalam *query*, namun mempunyai frekuensi yang tinggi. Cara ini menggunakan fungsi logaritmik matematika untuk memperoleh nilai tf.
- c. *Binary term frequency.*  
 Hanya memperhatikan apakah suatu *term* ada atau tidak dalam dokumen. Jika ada, maka tf diberi nilai 1, jika tidak ada diberi nilai 0. Pada cara ini jumlah kemunculan *term* dalam dokumen tidak berpengaruh.
- d. *Augmented term frequency.*


**Hak Cipta Dilindungi Undang-Undang**

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

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

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

Nilai  $tf$  adalah jumlah kemunculan suatu *term* pada sebuah dokumen, sedangkan nilai  $max(tf)$  adalah jumlah kemunculan terbanyak sebuah *term* pada dokumen yang sama.

#### 2.4.5.2 Inverse Document Frequency

*Inverse document frequency* (IDF) digunakan untuk memberikan tekanan terhadap dominasi *term* yang sering muncul di berbagai dokumen. Hal ini diperlukan karena *term* yang banyak muncul di berbagai dokumen, dapat dianggap sebagai *term* umum (*common term*) sehingga tidak penting nilainya. Pembobotan akan memperhitungkan faktor kebalikan frekuensi dokumen yang mengandung suatu *term* (*inverse dokumen frequency*).

#### 2.4.6 Algoritma Enhanced Confix Stripping

*Enhanced Confix Stripping* adalah algoritma *stemming* untuk bahasa Indonesia. Algoritma ini merupakan perbaikan terhadap algoritma *confix stripping stemmer* yang mana algoritma ini juga merupakan perbaikan terhadap algoritma nazief dan andriani (Tahitoe & Purwitasari, 2010) *Enhanced Confix Stripping (ECS) Stemmer* adalah Algoritma *stemming* kata pada Bahasa Indonesia dengan performa yang paling baik (memiliki jenis kesalahan *stemming* yang paling sedikit) (Arifin dkk, 2009). Algoritma ini menggunakan kamus kata dasar dan mendukung *recoding*, yakni penyusunan kembali kata-kata yang mengalami proses *stemming* berlebih.

Untuk memperbaiki kekurangan dalam algoritma *confix stripping stemmer*, Algoritma *enhanced confix stripping* menambah aturan pemenggalan dan membuat aturan untuk mengatasi kesalahan pemenggalan akhiran yang seharusnya tidak dilakukan. Aturan algoritma ini disebut *loop Pengembalian Akhiran*, dan ini akan dilakukan apabila proses *recoding* gagal. Aturan algoritma *loop pengembalian akhiran* dideskripsikan sebagai berikut:

1. Kembalikan seluruh awalan yang telah dihilangkan sebelumnya, sehingga menghasilkan model kata seperti berikut:  $[DP+[DP+[DP]]] + \text{Kata Dasar}$ .



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

Pemenggalan awalan dilanjutkan dengan proses pencarian di kamus kemudian dilakukan pada kata yang telah dikembalikan menjadi model tersebut.

2. Kembalikan akhiran sesuai dengan urutan model pada bahasa Indonesia. Ini berarti bahwa pengembalian dimulai dari DS (“-i”, “-kan”, “-an”), lalu PP (“-ku”, “-mu”, “-nya”), dan terakhir adalah P (“-lah”, “-kah”, “-tah”, “-pun”). Untuk setiap pengembalian, lakukan langkah 3) hingga 5) berikut. Khusus untuk akhiran “-kan”, pengembalian pertama dimulai dengan “k”, baru kemudian dilanjutkan dengan “an”.
3. Lakukan pengecekan di kamus kata dasar. Apabila ditemukan, proses dihentikan. Apabila gagal, maka lakukan proses pemenggalan awalan berdasarkan aturan pada Tabel 2.1.
4. Lakukan *recoding* apabila diperlukan.
5. Apabila pengecekan di kamus kata dasar tetap gagal setelah *recoding*, maka awalan-awalan yang telah dihilangkan dikembalikan lagi.

Adapun keterangan simbol huruf dari proses diatas sebagai berikut:

- PP = kata ganti positif, misalnya “-ku”, “-mu”, dan “-nya”
- DS = Asal mula kata akhiran, misalnya “-i”, “-kan”, dan “-an”.
- DP = Asal mula kata awalan, aturan kata awalan yang diterapkan secara langsung ke kata dasar atau ke kata-kata yang memiliki hingga dua asal mula kata awalan lain. Jenis awalan ini dibagi menjadi awalan rumit (“me-”, “be-”, “pe-”, dan “te”) dan awalan sederhana (“di-”, “ke-” dan “se-”).

Beberapa aturan dalam melakukan proses *stemming* pada algoritma *enhanced confix stripping* (Tahitoe & Purwitasari, 2010).

**Tabel 2.1 Aturan Pemenggalan Imbuhan Algoritma Enhanced Confix Stripping**

Aturan	Format Kata	Pemenggalan
1	berV...	ber-V...   be-rV...
2	berCAP...	ber-CAP... dimana C != 'r' & P != 'er'





**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.

Aturan	Format Kata	Pemenggalan
3	berCAerV...	ber-CaerV... dimana C!=’r’
4	belajar	bel-ajar
5	beC1erC2...	be-C1erC2... dimana C1!={’r’   ’l’}
6	terV...	ter-V...   te-rV...
7	terCerV...	ter-CerV... dimana C!=’r’
8	terCP...	ter-CP... dimana C!=’r’ dan P!=’er’
9	teC1erC2...	te-C1erC2... dimana C1!=’r’
10	me{l r w y}V...	me-{l r w y}V...
11	mem{b f v}...	mem-{b f v}...
12	mempe...	mem-pe...
13	mem{rV V}...	me-m{rV V}...   me-p{rV V}...
14	men{c d j s z}...	men-{c d j s z}...
15	menV...	me-nV...   me-tV
16	meng{g h q k}...	meng-{g h q k}...
17	mengV...	meng-V...   meng-kV   (mengV jika.....V=’e’)
18	menyV...	meny-sV...
19	mempA...	mem-pA... dengan A!=’e’
20	mempA...	mem-pA... dengan A!=’e’
21	perV...	per-V...   pe-rV...
22	perCAP	per-CAP... dimana C!=’r’ dan P!=’er’
23	perCAP	per-CAP... dimana C!=’r’ dan P!=’er’
24	perCAerV...	per-CAerV... dimana C!=’r’
25	pem{b f V}...	pem-{b f V}...



- 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.

Aturan	Format Kata	Pemenggalan
26	$\text{pem}\{rV V\}...$	$\text{pe-m}\{rV V\}...   \text{pe-p}\{rV V\}...$

Keterangan simbol huruf dari tabel diatas,

- C = huruf konsonan
- V = huruf vokal
- A = huruf vokal atau konsonan
- P = partikel atau fragmen dari suatu kata, misalnya “er”

### 2.4.7 Algoritma Latent Sematic Indexing

Keuntungan menggunakan model LSI adalah model ini *fully automatic* dan tidak menggunakan language expertise, akibatnya panjang vektor dokumen menjadi lebih pendek. Dengan membandingkan Model LSI dengan *multidimensional scaling*, ternyata ruang dokumen (document space) dengan model LSI lebih optimal pada saat menggunakan *inner product similarity function* demikian juga dengan menggunakan ukuran kesamaan lainnya.

*Latent Semantic Indexing* (LSI) adalah metode matematika dan statistika untuk menentukan hubungan kontekstual arti kata pada bagian teks yang dibutuhkan (Landauer, Foltz and Laham, 1998). Pada LSI dilakukan beberapa proses, yaitu:

1. *Singular Value Decomposition* (SVD) adalah representasi komponen kata dan dokument kedalam bentuk matrik.
2. *Cosine Similarity* dikenal sebagai rumus umum untuk mengukur kemiripan kata atau dokumen.

#### 2.1.1.1 Singular Value Decomposition

*Singular Value Decomposition* (SVD) merupakan representasi komponen kata dan dokumen kedalam bentuk matrik, sehingga dari hasil dekomposisi SVD terdapat tiga jenis operasi perbandingan yang dapat dilakukan, diantaranya yaitu (Aji et al., 2011)

1. Membandingkan dua kata (*terms*)

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.

Koordinat dari suatu kata pada *semantic space* direpresentasikan oleh vektor baris dari matrik  $U \times S$  yang bersesuaian dengan kata tersebut. Untuk menentukan kemiripan antara dua kata menggunakan metode *cosine similarity* dengan menghitung sudut antara kedua koordinat kata. Operasi inilah yang digunakan pada tugas akhir ini.

2. Membandingkan dua dokumen (*documents*)

Koordinat dari suatu dokumen pada *semantic space* direpresentasikan oleh vektor baris dari matrik  $V \times S$  yang bersesuaian dengan dokumen tersebut. Untuk menentukan kemiripan antara dua dokumen menggunakan metode *cosine similarity* dengan menghitung sudut antara kedua koordinat dokumen.

3. Membandingkan kata (*term*) dengan dokumen (*document*)

Berbeda dari dua operasi diatas yang memerlukan *cosine similarity* untuk menghitung kesamaan kata atau dokumen, untuk membandingkan kata  $i$  dengan dokumen  $j$  dapat diketahui dari nilai cell  $(i, j)$  dari matrik aproksimasi kata-dokumen yang didapat dari perhitungan SVD.

Menurut (Aji et al., 2011) SVD akan menguraikan sebuah matriks menjadi tiga buah matriks baru yaitu matriks vektor singular kiri, matriks nilai singular dan matriks vektor singular kanan, SVD dari sebuah matriks  $A$  berdimensi  $m \times n$  adalah sebagai berikut:

$$A_{m \times n} = U_{m \times m} \times S_{m \times n} \times V^T_{n \times n} \dots\dots\dots(2.3)$$

Keterangan:

$A$  : matriks berdimensi  $m \times n$

$U$  : matriks vektor singular kiri berdimensi  $m \times m$

$S$  : matriks nilai singular berdimensi  $m \times n$  dengan nilai terurut menurun

$V$  : matriks vektor singular kanan berdimensi  $n \times n$

$V^T$  : matrik  $V$  transpose.

Untuk lebih memahami cara memperoleh matrik-matrik tersebut yang diperoleh dari SVD, maka diberikan contoh sebagai berikut (Aji et al., 2011):



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

Misalkan diketahui matriks  $A_{2 \times 3}$  dengan nilai  $\begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix}$  akan

dikomposisi menggunakan SVD. Pertama kali yang dilakukan adalah mencari matriks ortogonal  $U$ .

**1. Mencari Matriks  $U$**

Untuk mencari  $U$  dilakukan perkalian matriks dengan matriks *transpose*  $A$ .

$$AA^T = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix} \begin{bmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 11 & 1 \\ 1 & 11 \end{bmatrix}$$

Selanjutnya dicari *eigenvalues* dan *eigenvectors* dari  $AA^T$ . Untuk mendapatkan *eigenvectors* digunakan rumus  $A\vec{v} = \lambda\vec{v}$  pada  $AA^T$

$$\begin{bmatrix} 11 & 1 \\ 1 & 11 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \lambda \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

$$11x_1 + x_2 = \lambda x_1 \rightarrow 11x_1 + x_2 - \lambda x_1 = 0 \rightarrow (11 - \lambda)x_1 + x_2 = 0$$

$$x_1 + 11x_2 = \lambda x_2 \rightarrow x_1 + 11x_2 - \lambda x_2 = 0 \rightarrow x_1 + (11 - \lambda)x_2 = 0$$

Cari nilai  $\lambda$  (*eigenvalues*) dengan menghitung determinan sama dengan nol ( $D=0$ ).

$$\begin{vmatrix} (11 - \lambda) & 1 \\ 1 & (11 - \lambda) \end{vmatrix} = 0$$

$$(11 - \lambda)(11 - \lambda) - 1 * 1 = 0$$

$$\lambda^2 - 22\lambda + 120 = 0$$

$$(\lambda - 10)(\lambda - 12) = 0$$

$$\lambda = 10, \lambda = 12$$

Substitusi  $\lambda$  pada persamaan sehingga didapat

**Tabel 2. 2 Substitusi  $\lambda$**

$(\lambda = 12)$	$(\lambda = 10)$
$(11 - \lambda) X_1 + X_2 = 0$	$(11 - \lambda) X_1 + X_2 = 0$
$(11 - 12) X_1 + X_2 = 0$	$(11 - 10) X_1 + X_2 = 0$
$X_1 = X_2$	$X_1 = X_2$





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.

Dari persamaan di atas didapatkan *eigenvectors*  $\lambda = 12$  adalah  $[1,1]$ , dan  $\lambda = 10$  adalah  $[1,-1]$ . Kemudian untuk mendapatkan matriks dari *eigenvectors* dilakukan dengan memposisikan *eigenvectors* tersebut terurut *descending* (menurun) dari kolom pertama ke kolom selanjutnya. Dengan demikian maka *eigenvectors* untuk adalah kolom pertama, dan  $\lambda = 10$  adalah kolom kedua, dapat dituliskan sebagai berikut:

$$\begin{matrix} \text{Kolom 1} & \text{kolom 2} \\ \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \end{matrix}$$

Selanjutnya matriks yang didapat ini dikonversi menjadi matriks orthogonal dengan menggunakan proses Gram-Schmidt *orthonormalization* pada vektor kolom.

Dimulai dari:  $\vec{v}_1$

$$\vec{u}_1 = \frac{\vec{v}_1}{|\vec{v}_1|} = \frac{[1,1]}{\sqrt{1^2 + 1^2}} = \frac{[1,1]}{\sqrt{2}} = \left[ \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right]$$

Untuk mendapatkan  $\vec{u}_2$  dilakukan dahulu perhitungan  $\vec{w}_2$  Rumus umum untuk menghitung  $\vec{w}_k$  adalah

$$\vec{w}_k = \vec{v}_k - \sum_{i=1}^{k-1} \vec{u}_i \cdot \vec{v}_k * \vec{u}_i$$

$$\vec{w}_2 = \vec{v}_2 - \vec{u}_1 * \vec{v}_2 * \vec{u}_1$$

$$\vec{w}_2 = [1, -1] - \left[ \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right] * [1, -1] * \left[ \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right]$$

Sehingga didapatkan matriks:

$$\mathbf{u} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

**2. Mencari Matriks  $V^T$**

$$\vec{u}_2 = \frac{\vec{v}_2}{|\vec{v}_2|} = \frac{[1,-1]}{\sqrt{1^2 + (-1)^2}} = \frac{[1,-1]}{\sqrt{2}} = \left[ \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \right]$$

$$\vec{w}_2 = [1, -1] - 0 * \left[ \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right]$$

$$\vec{w}_2 = [1, -1] - [0,0]$$

$$\vec{w}_2 = [1, -1]$$



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

Pencarian  $V^T$  mirip dengan pencarian matriks U. Namun dimulai dengan mengalikan matriks *transpose* A dengan matriks A.

$$A^T A = \begin{bmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 10 & 0 & 2 \\ 0 & 10 & 4 \\ 2 & 4 & 2 \end{bmatrix}$$

Selanjutnya mencari *eigenvalues* dan *eigenvectors* dari  $A^T A$ . untuk mendapatkan *eigenvectors* digunakan rumus  $A\vec{v} = \lambda\vec{v}$  pada  $A^T A$ .

$$\begin{bmatrix} 10 & 0 & 2 \\ 0 & 10 & 4 \\ 2 & 4 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$10x_1 + 2x_3 = \lambda x_1 \quad \Rightarrow (10 - \lambda)x_1 + 2x_3 = 0$$

$$10x_2 + 4x_3 = \lambda x_2 \quad \Rightarrow (10 - \lambda)x_2 + 4x_3 = 0$$

$$2x_1 + 4x_2 + 2x_3 = \lambda x_3 \quad \Rightarrow 2x_1 + 4x_2 + (2 - \lambda)x_3 = 0$$

Cari nilai  $\lambda$  (*eigenvalues*) dengan menghitung determinan sama dengan nol ( $D=0$ )

$$\begin{vmatrix} (10 - \lambda) & 0 & 2 \\ 0 & (10 - \lambda) & 4 \\ 2 & 4 & (2 - \lambda) \end{vmatrix} = 0$$

$$(10 - \lambda) \begin{vmatrix} (10 - \lambda) & 4 \\ 4 & (2 - \lambda) \end{vmatrix} - 0 \begin{vmatrix} 0 & 4 \\ 2 & (2 - \lambda) \end{vmatrix} + 2 \begin{vmatrix} 0 & (10 - \lambda) \\ 2 & 4 \end{vmatrix} = 0$$

$$(10 - \lambda) [(10 - \lambda)(2 - \lambda) - 16] - 0 + 2 [0 - (20 - 2\lambda)] = 0$$

$$\lambda(\lambda - 10)(\lambda - 12) = 0$$

$\lambda = 0, \lambda = 10, \lambda = 12$  Substitusi  $\lambda$  pada persamaan sehingga didapat: 1

**Tabel 2. 3 Matriks  $V^T$  – Substitusi  $\lambda$**

$\lambda = 12$	$\lambda = 10$	$\lambda = 0$
$(10 - 12)x_1 + 2x_3 = 0$	$(10 - 10)x_1 + 2x_3 = 0$	$10x_1 + 2x_3 = 0$
$-2x_1 + 2x_3 = 0$	$2x_3 = 0$	$x_3 = -5$
$x_1 = 1, x_3 = 1$	$x_3 = 0$	$10x_2 + 20 = 0$
$(10 - 12)x_2 + 4x_3 = 0$	$2x_1 + 4x_2 = 0$	$x_2 = 2$
$-2x_2 + 4x_3 = 0$	$x_1 = -2x_2$	$2x_1 + 8 + 10 = 0$
$x_2 = 2x_3, x_2 = 2$	$x_1 = 2, x_2 = -1$	$x_1 = 1$

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.

Dengan demikian maka *eigenvectors* untuk  $\lambda = 12$  adalah kolom pertama,  $\lambda = 10$  adalah kolom kedua, dan  $\lambda = 0$  adalah kolom ketiga. dapat dituliskan sebagai berikut:

Kolom 1	Kolom 2	Kolom 3
1	2	1
2	-1	2
1	0	-5

Selanjutnya matriks yang didapat ini dikonversi menjadi matriks orthogonal dengan menggunakan proses Gram-Schmidt *orthonormalization* pada vektor kolom

Dimulai dari  $\vec{v}_1$

$$\vec{u}_1 = \frac{\vec{v}_1}{|\vec{v}_1|} = \left[ \frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}} \right]$$

Untuk mendapatkan  $\vec{u}_2$  dilakukan dahulu perhitungan  $\vec{w}_2$

$$\vec{w}_2 = \vec{v}_2 - \vec{u}_1 * \vec{v}_2 * \vec{u}_1 = [2, -1, 0]$$

$$\vec{u}_2 = \frac{\vec{v}_2}{|\vec{v}_2|} = \left[ \frac{2}{\sqrt{5}}, \frac{-1}{\sqrt{5}}, 0 \right]$$

Untuk mendapatkan dilakukan dahulu perhitungan

$$\vec{w}_3 = \vec{v}_3 - \vec{u}_1 * \vec{v}_3 * \vec{u}_1 = \left[ \frac{-2}{3}, \frac{-4}{3}, \frac{10}{3} \right]$$

$$\vec{u}_3 = \frac{\vec{v}_3}{|\vec{v}_3|} = \left[ \frac{1}{\sqrt{30}}, \frac{2}{\sqrt{30}}, \frac{-5}{\sqrt{30}} \right]$$

Sehingga didapatkan matriks:

$$V = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{30}} \\ \frac{2}{\sqrt{6}} & \frac{-1}{\sqrt{5}} & \frac{2}{\sqrt{30}} \\ \frac{1}{\sqrt{6}} & 0 & \frac{-5}{\sqrt{30}} \end{bmatrix}$$

Namun yang dibutuhkan adalah  $V^T$ , maka dilakukan transpose terhadap V.

$$V^T = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{30}} \\ \frac{2}{\sqrt{6}} & \frac{-1}{\sqrt{5}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{30}} & \frac{-5}{\sqrt{30}} \end{bmatrix}$$



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.

### 3. Mencari Matriks S (Σ)

Matriks S (Σ) didapatkan dari nilai akar *eigenvalues* yang bukan nol. *Eigenvalues* tersebut disusun secara diagonal dari yang paling besar mulai dari S<sub>11</sub>, S<sub>22</sub>, S<sub>33</sub> dan seterusnya sampai *eigenvalues* paling kecil, S<sub>mm</sub>. Kemudian agar bisa dikalikan dengan matriks U dan V<sup>T</sup> pada rumus SVD maka perlu ditambahkan kolom yang berisi nol sehingga matriksnya menjadi sebagai berikut:

$$S = \begin{bmatrix} \sqrt{12} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{bmatrix}$$

Dari matriks-matriks hasil dekomposisi akan dapat dihitung kedekatan antara dokumen dengan menggunakan *cosine Similarity*.

#### 2.3.5.2 *Cosine Similarity*

*Cosine Similarity* dikenal sebagai rumus yang umumnya digunakan untuk pengukuran similaritas, dengan menentukan sudut antara vektor dokumen dengan vektor kueri dalam dimensi V ruang Euclidean, dengan V adalah ukuran *input* kata (Lee, Chuang and Seamons, 1997). Hasil *cosine similarity* bernilai antara 0 sampai 1. Nilai 0 menunjukkan bahwa dokumen tidak terkait atau tidak berhubungan dengan kueri. Dan jika hasil *cosine similarity* bernilai 1, berarti keterhubungan antara dokumen dengan kueri tinggi. Dengan rumus sebagai berikut:

**Vektor kueri:**

$$\bar{q} = q^T U_k \Sigma_k^{-1} \dots \dots \dots (2.4)$$

**Keterangan :**

$\bar{q}$  : Vektor kueri

$q^T$  : Transpose matrik vektor kueri

$U_k$  : Matrik Singular kiri dalam ruang k

$\Sigma_k^{-1}$  : Invers Matrik Singular dalam ruang k





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.

**Vektor dokumen:**

Pembentukan vektor dokumen ( $\bar{d}$ ) dilakukan setelah proses SVD. vektor dokumen sama dengan setiap kolom pada matrik  $V^T$ .

Setelah vektor dokumen dan vektor kueri dibentuk maka dihitunglah *similarity* antara kueri dan dokumen dengan cara menghitung nilai cosinus sudut yang dibentuk oleh vektor kueri dan vektor dokumen. *Similarity* vektor kueri dan vektor dokumen didefinisikan sebagai berikut (Aji dkk, 2011) :

$$\text{cosine similarity} = \frac{\bar{q} \cdot \bar{d}}{\|\bar{q}\| \|\bar{d}\|} = \frac{\sum_{i=1}^n \bar{q}_i \times \bar{d}_i}{\sqrt{\sum_{i=1}^n (\bar{q}_i)^2 \times \sum_{i=1}^n (\bar{d}_i)^2}} \dots\dots\dots(2.5)$$

Keterangan :

- $\bar{q}$  : Vektor kueri
- $\bar{d}$  : Vektor dokumen
- $\bar{q} \cdot \bar{d}$  : Dot produk antara vektor kueri dan dokumen
- $\|\bar{q}\|$  : Panjang vektor kueri
- $\|\bar{d}\|$  : Panjang vektor dokumen
- $\|\bar{q}\| \|\bar{d}\|$  : Cross produk antara  $\|\bar{q}\|$  dan  $\|\bar{d}\|$

Menurut Deerwester (1990), ada beberapa keunggulan dari teori ini:

1. Sinonim, LSI ini mampu mengenali kata yang berbeda tapi memiliki kesamaan arti.
2. Kata yang memiliki banyak makna (Polisemi) akan mengurangi keakuratan dalam pencarian dokumen, teknik *Reduced Representation* pada LSI diharapkan akan dapat menghilangkan *noise* pada kata.
3. Keterkaitan kata, LSI memperhatikan keterkaitan kata dalam dokumen.

**2.5 Evaluasi Kualitas Temu Kembali**

Setelah melakukan pencarian dan mendapatkan hasilnya, suatu sistem harus diuji tingkat relevan dari masukan *query* dengan hasil yang didapatkan apakah sesuai dengan yang diinginkan oleh pengguna. Terdapat dua kategori dokumen yang



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.

dihasilkan oleh sistem IR terkait pemrosesan kueri yaitu dokumen yang relevan dan dokumen yang diterima oleh pengguna. Secara umum yang digunakan untuk mengukur kualitas hasil pencarian yaitu dengan menggunakan *precision*, *recall* Akurasi (Utami & Aliandu, 2013).

*Precision* adalah rasio jumlah dokumen relevan yang ditemukan dengan total jumlah dokumen yang ditemukan oleh sistem pencarian. *Precision* mengindikasikan kualitas himpunan jawaban, tetapi tidak memandang total jumlah dokumen yang relevan dalam kumpulan dokumen. Berikut ini adalah persamaan untuk mencari nilai

$$Precision = \frac{|{\text{relevant documents}} \cap {\text{documents retrieved}}|}{|{\text{documents retrieved}}|} \dots\dots\dots(2.6)$$

Sedangkan *Recall* adalah rasio jumlah dokumen relevan yang ditemukan kembali dengan total jumlah dokumen dalam kumpulan dokumen yang dianggap relevan. Berikut ini adalah persamaan untuk mencari nilai *Recall* :

$$Recall = \frac{|{\text{relevant documents}} \cap {\text{documents retrieved}}|}{|{\text{relevant documents}}|} \dots\dots\dots(2.7)$$

Kemudian akan juga dihirung akurasinya dimana akurasi adalah rasio jumlah dokumen relevan yang ditemukan kembali ditambah jumlah dokumen yang tidak relevan yang tidak ditemukan kembali dengan total jumlah seluruh dokumen.

Akurasi =

$$\frac{|{\text{relevant documents}} \cap {\text{documents retrieved}}| + |{\text{unrelevant documents}} \cap {\text{documents retrieved}}|}{|{\text{all documents}}|} \dots\dots\dots(2.8)$$

**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.

## 2.6 Penelitian Terkait

Berikut Tabel 2.4 Ini adalah penelitian terkait dari penelitian yang akan di lakukan :

**Tabel 2. 4 Penelitian Tekait**

No	Penulis	Tahun	Judul	Dipublikasi Pada	Kesimpulan
1	Wulandari	2013	Sistem Pencarian Ayat Al-Qur'an Berdasarkan Terjemahan Bahasa Indonesia dengan Pemodelan Ruang Vektor	Jurnal Sistem Informasi FST UIN SUSKA RIAU	Hampir memberikan hasil yang relevan oleh sistem yang dibangun, hal tersebut berdasarkan pengujian dengan memasukkan dua belas <i>Query</i> kedalam pencarian dengan perhitungan <i>precision</i> dan <i>recall</i>
2	Muhammad dkk	2013	<i>A Detailed Study on Information Retrieval using Genetic Algorithm</i>	Journal of Industrial and Intelligent Information Vol. 1, No. 3, September 2013	Menjelaskan bagaimana menentukan solusi dalam sistem temu kembali dengan menjelaskan bagaimana penerapan metode Latent Semantic Analisis pada sistem temu kembali dokumen



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	Penulis	Tahun	Judul	Dipublikasi Pada	Kesimpulan
3	Rasyidi dkk	2013	<i>Indonesian Hadith Retrieval System Using Thesaurus</i>	International Conference on Computer, Control, Informatics and Its Applications	Bahwa penggunaan tesaurus pada sistem temu kembali hadist juga dapat meningkatkan kualitas hasil pencarian dalam menemukan dokumen yang lebih relevan, terutama pada tingkat pengukuran <i>recall</i> nya dengan penggunaan <i>Query</i> yang lebih meluas.
4	kumar dkk	2012	<i>Analysis of a Vector Space Model, Latent Semantic Indexing and Formal Concept Analysis for Information Retrieval</i>	ASWANI KUMAR, Ch; RADVANSKY, M.; ANNAPURNA, J. <i>Analysis of a vector space model, latent semantic indexing and formal concept analysis for information retrieval. Cybernetics and Information Technologies, 2012, 12.1: 34-48.</i>	penelitian ini menjelaskan tentang kelebihan metode LSI dalam menangkap informasi yg tersembunyi dalam dokumen yang berbentuk matriks dimensi menggunakan SVD.





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

No	Penulis	Tahun	Judul	Dipublikasi Pada	Kesimpulan
5	Kumar	2009	<i>On the Performance of Latent Semantic Indexing-based Information Retrieval</i>	CHERUKURI, Aswani Kumar; SURIPEDDI, Srinivas. On the performance of latent semantic indexing based information retrieval. <i>CIT. Journal of Computing and Information Technology</i> , 2009, 17.3: 259-264.	penelitian ini menjelaskan pada model LSI pada IR lebih baik dari model VSM <i>classic</i> dengan rata-rata 30% lebih baik
6	Fadly Zadid Taqwa	2015	Penerapan Metode <i>Latent Semantic Analysis</i> pada Sistem Temu Kembali Informasi Hadits	20-05-2015	Penggunaan <i>thesaurus</i> untuk penentuan sinonim antara kueri dan dokumen dianggap baik karna mampu mengetahui maksud lain dari pengguna seperti yang telah dilakukan pada tahap pengujian.