

# 22\_PENGGUNAAN SISTEM TEMU KEMBALI DALAM PENCARIAN KATA

*by* Herbertus Yulianton

---

**Submission date:** 11-Apr-2023 01:08AM (UTC+0700)

**Submission ID:** 2060758576

**File name:** 22\_PENGGUNAAN\_SISTEM\_TEMU\_KEMBALI\_DALAM\_PENCARIAN\_KATA.pdf (357.09K)

**Word count:** 1717

**Character count:** 9974

## PENGUNAAN SISTEM TEMU KEMBALI DALAM PENCARIAN KATA UNTUK TERJEMAHAN AL QURAN

Sri Eniyati,<sup>1</sup>Rina Candra Noor Santi,<sup>2</sup>Heribertus Yulianton<sup>3</sup>

Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Stikubank  
e-mail: <sup>1</sup>eniyati03@gmail.com, <sup>2</sup>r\_candra\_ns@edu.unisbank.ac.id, <sup>3</sup>heri @edu.unisbank.ac.id

### ABSTRAK

Di jaman yang semakin canggih ini, teknologi informasi berkembang cukup pesat. Penyimpanan-penyimpanan file masih banyak dilakukan, sehingga dalam pencarian data akan sulit dilakukan. Dengan kondisi seperti ini, maka dikembangkanlah metode pencarian data yang dikenal dengan Sistem Informasi Temu Kembali atau Information Retrieval.

Information Retrieval yaitu ilmu pencarian informasi pada dokumen, pencarian untuk dokumen itu sendiri, mencari di dalam database untuk teks, suara, gambar, atau data lainnya. Vector Space Model merupakan model Information Retrieval yang merepresentasikan dokumen dan query dalam bentuk vektor dimensional. Konsep dasar dari VSM adalah menghitung jarak antar dokumen kemudian mengurutkan berdasarkan tingkat kedekatannya. Semakin kecil jarak antardokumen, maka semakin mirip keduanya.

Metode VSM dipilih karena cara kerja model ini efisien, mudah dalam representasi dan dapat diimplementasikan pada document-matching.

Keyword : Information Retrieval, Vector Space Model

### 1. PENDAHULUAN

Teknologi Informasi adalah suatu studi perancangan, implementasi, pengembangan, dukungan atau manajemen sistem informasi berbasis komputer, khususnya perangkat keras (*hardware*) dan perangkat lunak (*software*). Dengan kata lain, teknologi informasi adalah berbagai fasilitas yang terdiri dari hardware dan software untuk mendukung dan meningkatkan kualitas informasi bagi masyarakat dengan cepat dan berkualitas.

Pencarian kata merupakan salah satu bagian dari *Information Retrieval*. *Information Retrieval* merupakan bagian dari computer science yang berhubungan dengan pengambilan informasi dari dokumen-dokumen yang didasarkan pada isi dan konteks dari dokumen-dokumen itu sendiri. *Information Retrieval* yaitu ilmu pencarian informasi pada dokumen, pencarian untuk dokumen itu sendiri, mencari di dalam database untuk teks, suara, gambar, atau data lainnya.

Ada tiga model yang digunakan dalam *information retrieval*, yang pertama *Probabilistic model*, contoh model ini ialah penerapan teorema Bayes dalam model *probabilistic*, yang kedua *Set-theoretic models*, contoh model ini ialah *Standard boolean* dan yang terakhir *Algebraic model*, contoh model ini adalah merepresentasikan dokumen dan query sebagai *vector similarity* antara *vector* dokumen dan *vector query*. Contoh model ini ialah *Vector space model*. Dari tiga model *information retrieval*, *Algebraic model* dengan contoh model *Vector space model* adalah model yang paling sederhana dalam pencarian kata, telah terbukti memiliki efektifitas dalam pencarian kata dengan menampilkan hasil pencariannya berdasar kemiripan *vector query* dan *vector* dokumen.

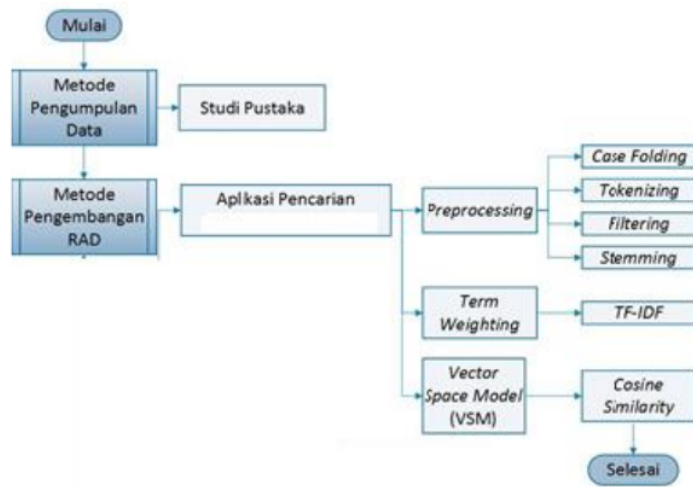
*Vector Space Model* merupakan model IR yang merepresentasikan dokumen dan query dalam bentuk vektor dimensional. Konsep dasar dari VSM adalah menghitung jarak antar dokumen kemudian mengurutkan berdasarkan tingkat kedekatannya. Semakin kecil jarak antardokumen, maka semakin mirip keduanya (Bari & Saputra, 2011, hal. 3-4).

Metode VSM dipilih karena cara kerja model ini efisien, mudah dalam representasi dan dapat diimplementasikan pada document-matching. Berdasarkan keterangan tersebut maka penulis menggunakan Pendekatan TF-IDF pada Vector Space Model (VSM) sebagai model pencarian kata dalam terjemahan Al-Qur'an.

### 2. Metodologi Penelitian

#### 2.1. Kerangka Berfikir

Kerangka berfikir pada saat melakukan penelitian dapat dilihat dari gambar 1 dibawah ini.



Gambar 1. Kerangka Berfikir

**2.2. Metode Penelitian**

1. Metode Pengumpulan Data
  - a. Observasi  
Observasi dilakukan dengan mencari terjemahan Al Quran untuk mendapatkan data.
  - b. Studi Kepustakaan  
Studi kepustakaan dilakukan untuk memperoleh literatur-literatur yang relevan dengan obyek penelitian.
2. Dalam penelitian ini, metode yang digunakan untuk analisa dan perancangan meliputi :
  - a. Analisis  
Pada tahap ini dicari alternatif-alternatif pemecahan masalah yang paling sesuai untuk mengatasi permasalahan yang ada.
  - b. Desain  
Merancang pemecahan masalah untuk menentukan langkah-langkah operasi, prosedur, sekaligus membuat desain secara menyeluruh yang meliputi data dan prosedur dan algoritma pemrograman.
  - c. Implementasi  
Implementasi untuk menghitung menggunakan microsoft excel sesuai dengan kebutuhan yang ada.

**2.3. Jenis Penelitian**

Penelitian ini merupakan penelitian survei, yaitu penelitian yang mengambil sampel secara langsung dari populasi. Dilihat dari permasalahan yang diteliti, penelitian ini merupakan penelitian kausalitas, yang bertujuan untuk menganalisis dan mencari salah satu kata terjemahan di dalam Al Quran dengan surah yang di teliti adalah Surah Al-Fajr dengan 30 ayat.

**3. HASIL DAN PEMBAHASAN**

**3.1. Indikator yang digunakan**

Dari masing-masing indikator yang telah diturunkan kedalam indikator operasional, diperjelas kembali dalam variabel-variabel yang mudah untuk diukur secara kuantitatif. Berikut data yang diambil dapat ditunjukkan pada tabel 1.

Tabel 1. Beberapa Contoh Terjemahan Surah Al-Fajr

No	Terjemahan
1	Demi Fajar
2	Demi malam yang sepuluh
3	Demi yang genap dan ganjil
4	Demi malam apabila berlalu

5	Adakah pada yang demikian itu terdapat sumpah (yang dapat diterima) bagi orang-orang yang berakal?
6	Tidakkah engkau (Muhammad) memperhatikan bagaimana Tuhanmu berbuat terhadap (kaum) 'Ad?
7	(yaitu) penduduk Iram (ibukota kaum 'Ad) yang mempunyai bangunan-bangunan yang tinggi,
8	yang belum pernah dibangun (suatu kota) seperti itu di negeri-negeri lain,
9	dan (terhadap) kaum amud yang memotong batu-batu besar di lembah,
10	dan (terhadap) Fir'aun yang mempunyai pasak-pasak (bangunan yang besar),
11	yang berbuat sewenang-wenang dalam negeri,

**3.2. Analisa**

**3.2.1 Analisa**

1. Case Folding

*Case Folding* adalah tahapan pertama, yaitu menghilangkan karakter selain huruf dan mengubah semua *term* ke dalam huruf kecil.

Tabel. 2. Contoh Case Folding

No	sebelum case folding	sesudah case folding
1	Demi malam apabila berlalu	demi malam apabila berlalu
2	Adakah pada yang demikian itu terdapat sumpah (yang dapat diterima) bagi orang-orang yang berakal?	adakah pada yang demikian itu terdapat sumpah yang dapat diterima bagi orang-orang yang berakal

2. Tokenizing

*Tokenizing* adalah proses memecah setiap term ke dalam array yang dipisahkan oleh spasi.

Tabel. 3. Contoh Tokenizing

No	Sebelum <i>Tokenizing</i>	Sesudah <i>Tokenizing</i>
1	yang belum pernah dibangun suatu kota seperti itu di negeri - negeri lain	yang
		belum
		pernah
		dibangun
		suatu
		kota
		seperti
		itu
		di
		negeri
negeri		

		lain
2	dan terhadap Fir'aun yang mempunyai pasak-pasak bangunan yang besar	dan
		terhadap
		firaun
		yang
		mempunyai
		pasak
		pasak
		bangunan
		yang
		besar

3. Filtering dan Stemming

*Filtering* adalah proses menghapus kata yang terdapat dalam *database stoplist*. Contoh kata yang dihapus adalah yang, dan, pada, itu, ini lain, dan sebagainya. Sedangkan *Stemming* adalah tahapan yang membuat sebuah term menjadi bentuk kata dasarnya. Contohnya adalah kalimat “mempunyai” menjadi kata dasar “punya”, kalimat “perkiraan” menjadi kata dasar “kira”, dan lain sebagainya.

3.2.3. Menghitung Term Weighting & Vector Space Model (VSM)

Dalam penelitian ini kata yang ingin dicari adalah kata “Tuhan” pada terjemahan Surah Al-Fajr, dengan langkah-langkahnya sebagai berikut:

1. Menghitung Term Frequency (TF-DF) dan Menghitung Document Frequency dan Inverse Document Frequency (IDF)

Dari hasil *stemming*, dilakukan penghitungan TF (*term frequency*) dan DF (*document frequency*). Yaitu TF merupakan jumlah suatu term pada setiap dokumen, dan DF adalah jumlah dokumen yang memiliki suatu term. Setelah didapatkan hasil perhitungan DF dilanjutkan dengan perhitungan IDF. Dalam perhitungan ini dibutuhkan nilai N yaitu total jumlah ayat yang digunakan yaitu 30 ayat. Berikut contoh perhitung TFIDF dapat dilihat pada tabel 4.

Tabel 4. Contoh Perhitungan (TF- DF) dan IDF

Token	Q	TF&DF					df	D/df	Idf
		D1	D2	D3	D4	D5			
Tuhan	1						9	3.333333	0.522879
Demi		1	1	1	1	1	5	6	0.778151
Fajar		1					1	30	1.477121
Malam			1		1		2	15	1.176091
Sepuluh			1				1	30	1.477121
Genap				1			1	30	1.477121
Ganjil				1			1	30	1.477121
Sumpah						1	1	30	1.477121
Terima						1	1	30	1.477121

orang						1	1	30	1.477121
akal						1	1	30	1.477121
buat							1	30	1.477121

## 2. Menghitung TF-IDF

Setelah didapatkan hasil perhitungan IDF dilanjutkan dengan perhitungan TF-IDF. Dalam perhitungan ini adalah mengkalikan hasil TF dengan hasil IDF, yang dapat dilihat pada tabel 5.

Tabel 5. Contoh Perhitungan (TF- DF) dan IDF

Token	Q	TF			df	D/df	Idf	W		
		D1	D2	D3				Q	D1	D2
Tuhan	1				9	3.333333	0.522879	0.522879	0	0
Demi		1	1	1	5	6	0.778151	0	0.778151	0.778151
Fajar		1			1	30	1.477121	0	1.477121	0
Malam			1		2	15	1.176091	0	0	1.176091
Sepuluh			1		1	30	1.477121	0	0	1.477121
Genap				1	1	30	1.477121	0	0	0
Ganjil				1	1	30	1.477121	0	0	0
Sumpah					1	30	1.477121	0	0	0
Terima					1	30	1.477121	0	0	0
Orang					1	30	1.477121	0	0	0
Akal					1	30	1.477121	0	0	0

## 3. Menghitung Cosine Similarity

Dalam proses ini, adalah mencari nilai cosine similarity. Adapun contoh perhitungan dapat dilihat pada tabel 6.

Tabel 6. Contoh Hasil Perhitungan cosine similarity

D13	D14	D15	D16	D17	D18	D19	D20	D21	D22
0.074749	0.074749	0.298995	0.298995	0	0	0	0	0	0.074749
0.00272	0.005057	0.011029	0.003134	0	0	0	0	0	0.001269
0.271977	0.505683	1.102855	0.313432	0	0	0	0	0	0.126871

## 4. KESIMPULAN

Kesimpulan yang dapat diberikan adalah sebagai berikut:

1. Dari hasil penelitian penulis menyimpulkan bahwa metode VSM untuk pencarian kata untuk mencari kata "Tuhan" pada Al Quran melalui tahapan pertama yaitu proses *text preprocessing* yang menggunakan 4 tahapan yaitu *case folding*, *tokenizing*, *filtering*, *stemming* (algoritma porter stemmer) dan *term weighting* (metode TF-IDF) yang berfungsi untuk memaksimalkan hasil pencarian. Dan proses similaritas (*cosine similarity*) yang berfungsi untuk mendapatkan kata yang dimaksud dengan query serta jaraknya untuk pengurutan.
2. Dari hasil maka didapatkan pada ayat 13, 14, 15, 16 dan 22 yang menggunakan kata "Tuhan" dalam terjemahan Al Quran.

**Daftar Pustaka**

- [1] Bari, A., & Saputra, R. H. (2011). *Penerapan Pencarian Kata dengan Vector Space Model pada Aplikasi Terjemahan Juz Amma berbasis Java ME*. Palembang: Prodi Teknik Informatika, STMIK GI MDP.
- [2] Brata, D., & Hetami, A. (2015). *Perancangan Information Retrieval (IR) Untuk Pencarian Ide Pokok Teks Artikel Berbahasa Inggris Dengan Pembobotan Vector Space Model*. Stmik Asia Malang.
- [3] Fitri, M. (2013). *Perancangan Sistem Temu Balik Informasi dengan Metode Pembobotan Kombinasi TF-IDF untuk Pencarian Dokumen Berbahasa Indonesia*. Pontianak: Prodi Teknik Informatika, Jurusan Teknik Elektro Fakultas Teknik, Universitas Tanjungpura.
- [4] Fauziah Tri M (2017). Implementasi Pendekatan VSM Pada Pencarian Terjemahan AlQuran Juz 1-13 Berbasis Web, Prodi Teknik Informatika, UIN Jakarta
- [5] Ingwersen, P, 1992, *Information retrieval Interaction, London, Taylor Graham Publishing*. <http://www.db.dk/pi/iri> [29 Agustus 2005]
- [6] Harjanto, D. S., Endah, S. N., & Bahtiar, N. (2012). *Sistem Temu Kembali Informasi pada Dokumen Teks Menggunakan Metode Term Frequency Inverse Document Frequency (TF-IDF)*. Semarang: Universitas Diponegoro, Fakultas Sains dan Matematika, Jurusan Matematika & Jurusan Ilmu Komputer/Informatika.

# 22\_PENGGUNAAN SISTEM TEMU KEMBALI DALAM PENCARIAN KATA

---

## ORIGINALITY REPORT

---

**23%**

SIMILARITY INDEX

**24%**

INTERNET SOURCES

**10%**

PUBLICATIONS

**18%**

STUDENT PAPERS

---

## MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

---

4%

★ Submitted to Universitas Stikubank

Student Paper

---

Exclude quotes  On

Exclude matches  < 2%

Exclude bibliography  On