Jurnal Antartika

STT Atlas Nusantara

Volume 08/2018/01

Elektronik Learning Pepak Bahasa Jawa Berbasis Android
Affan Mifbakhur Riza, Rizza Muhammad Arief, Irsyad Arif Mashudi

Pemanfaatan Text Mining dalam Pencarian Ayat AlQuran menggunakan TF-IDF dan Cosine Similarity
Nuzul Hikmah

Rancang Bangun Website Pembelajaran Berbasis Media Sosial Untuk Fitur Ujian Online Menggunakan Metode User Centered Design (UCD)
Revanda Retno Widuri Cahyaningrum, Betty Dewi Puspasri, Siyamta

Aplikasi Sistem Pakar Diagnosa Penyakit Ginjal Dengan Metode Dempster Shafer
Pramadhana Cahaya Kurnia, Alun Sujjada, Pieter Stephanus

Perancangan Sistem Open Access Journal Menggunakan Open Journal System
Irsyad Arif Mashudi, Yohan Pribadi

Sistem Pakar Diagnosa Penyakit Demam Berdarah Menggunakan Metode Naive Bayes dan Certainty Factor
Intan Shovi Anggra Winnie

informatic Engineering

ISSN: 2089-2837

JURNAL ANTARTIKA

Jurnal ATLAS NUSANTARA

Teknik Informatika

Volume 8, Nomor 1, April 2018

Elektronik Learning Pepak Bahasa Jawa Berbasis Android
Affan Mifbakhur Riza, Rizza Muhammad Arief

Pemanfaatan Text Mining dalam Pencarian Ayat AlQuran menggunakan TF-IDF dan Cosine Similarity
Nuzul Hikmah

Rancang Bangun Website Pembelajaran Berbasis Media Sosial Untuk Fitur Ujian Online Menggunakan Metode User Centered Design (UCD)
Revanda Retno Widuri Cahyaningrum, Betty Dewi Puspasri, Siyamta

Aplikasi Sistem Pakar Diagnosa Penyakit Ginjal Dengan Metode Dempster Shafer
Pramadhana Cahaya Kurnia, Alun Sujjada, Pieter Stephanus

Perancangan Sistem Open Acccess Journal Menggunakan Open Journal System
Irsyad Arif Mashudi, Yohan Pribadi

Sistem Pakar Diagnosa Penyakit Demam Berdarah Menggunakan Metode Naive Bayes Dan Certainty Factor
Intan Shovi Anggra Winnie

STT ATLAS NUSANTARA MALANG

PROGRAM STUDI TEKNIK INFORMATIKA

ANTARTIKA	Volume	Nomor	Halaman	Malang	ISSN
	8	1	1-61	April 2018	2089-2837

JURNAL ANTARTIKA

Volume 8, nomor 1, April 2018

PENERBIT (PUBLISHER)
ATLAS NUSANTARA ENGINEERING COLLEGE PRESS

ALAMAT PENYUNTING (EDITORIAL ADDRESS)
Jl. Teluk Pacitan No. 50 lantai 1 Arjosari Malang 65126
Telp: (0341) 475898 | Fax: (0341) 475897
Email: info@sttar.ac.id
Homepage: http://www.sttar.ac.id

Jurnal Antartika STTAR terbit sejak 2011 merupakan jurnal ilmiah sebagai bentuk pengabdian dalam pengembangan bidang Teknik Informatika dan bidang terkait lainnya.

Jurnal Antartika STTAR diterbitkan oleh jurusan Teknik Informatika STT Atlas Nusantara. Redaksi mengundang para profesional dari dunia usaha, pendidikan dan peneliti untuk menulis mengenai perkembangan ilmu bidang yang berkaitan dengan teknik Informatika.

Jurnal Antartika STTAR diterbitkan 2 (dua) kali dalam 1 tahun pada bulan April dan Oktober

JURNAL ANTARTIKA

Volume 8, nomor 1

DAFTAR ISI

1. **Elektronik Learning Pepak Bahasa Jawa Berbasis Android**
   Affan Mifbakhur Riza, Rizza Muhammad Arief 1-12

2. **Pemanfaatan Text Mining dalam Pencarian Ayat AlQuran menggunakan TF-IDF dan Cosine Similarity**
   Nuzul Hikmah 13-22

3. **Rancang Bangun Website Pembelajaran Berbasis Media Sosial Untuk Fitur Ujian Online Menggunakan Metode User Centered Design (UCD)**
   Revanda Retno Widuri Cahyaningrum, Betty Dewi Puspasri, Siyamta 23-32

4. **Aplikasi Sistem Pakar Diagnosa Penyakit Ginjal Dengan Metode Dempster Shafer**
   Pramadhana Cahaya Kurnia, Alun Sujjada, Pieter Stephanus 32-42

5. **Perancangan Sistem Open Acccess Journal Menggunakan Open Journal System**
   Irsyad Arif Mashudi, Yohan Pribadi 43-50

6. **Sistem Pakar Diagnosa Penyakit Demam Berdarah Menggunakan Metode Naive Bayes Dan Certainty Factor**
   Intan Shovi Anggra Winnie 51-61

Pemanfaatan Text Mining dalam Pencarian Ayat AlQuran menggunakan TF-IDF dan Cosine Similarity

Nuzul Hikmah
Teknik Elektro Konsentrasi Teknik Komputer, Universitas Panca Marga Probolinggo
n.hikmah1807@gmail.com

ABSTRAK

Text mining adalah suatu teknolgi penambangan data berupa teks. Tujuannya adalah untuk menemukan kata-kata yang bisa mewakili isi dari suatu dokumen sehingga dapat dicari keterhungan antar dokumen yang lain. Sebagian besar umat muslim masih belum memahami isi yang terkandung di dalam AlQuran. Penelitian ini memanfaatkan teknologi text mining dalam melakukan pencarian ayat AlQuran yang relevan dengan topik yang dicari. Algoritma yang digunakan yaitu pembobotan TF-IDF dan Cosine Similarity. Hasilnya adalah berupa pemograman Web dimana pengguna dapat memasukkan suatu topik yang hendak dicari di dalam AlQuran dan selanjutnya sistem akan memberikan output berupa urutan ayat AlQuran yang paling relevan dengan topik yang dicari.

Kata kunci : *Text Mining*, TF-IDF, *Cosine Similarity*

1. Pendahuluan

AlQuran merupakan kitab suci bagi umat Islam (Muslim). AlQuran merupakan wahyu dari Allah SWT yang diturunkan kepada Nabi Muhammad SAW. AlQuran mengandung nilai-nilai berhubungan dengan keimanan, akhlak, syariah serta peraturan-peraturan yang dapat mengatur tata cara hidup dan tingkah laku manusia, baik sebagai makhluk individu maupun makhluk sosial.[9] Beragamnya informasi yang terdapat di dalam AlQuran, tidak sedikit umat muslim yang merasa kesulitan untuk mencari suatu informasi tertentu.

Dengan adanya perkembangan teknologi yang bertambah besar maka kebutuhan untuk memberikan informasi yang cepat dan tepat menjadikan fokus utama dalam penelitian. Salah satunya dengan memanfaatkan teknologi text mining.

Text mining merupakan suatu penambangan informasi yang berguna dari data yang berupa tulisan. Text mining merupakan bagian dari data mining yang dapat memproses data dalam jumlah yang sangat besar. Dengan text mining akan dilakukan proses mencari atau penggalian informasi yang berguna dari data tekstual.[5] Tujuannya, agar dapat menemukan suatu informasi yang penting dalam suatu dokumen, dan mencari hubungan antara dokumen yang satu dengan dokumen yang lain.

Dalam penelitian ini, memanfaatkan text mining untuk pencarian ayat AlQuran dengan menggunakan metode pembobotan TF-IDF dan Cosine Similarity. Dengan harapan agar umat muslim tidak merasa kesulitan dalam mencari informasi yang terkandung di dalam AlQuran dengan cara memasukkan teks berupa topik yang dicari sehingga dapat menghasilkan urutan ayat AlQuran yang relevan dengan topik yang dicari tersebut.

2. Tinjauan Pustaka

2.1 Text Mining

Data mining merupakan satu cabang disiplin ilmu yang dapat menjelajahi dataset dalam ukuran yang sangat besar untuk dapat mengekstraksi informasi tersirat didalamnya, yang sebelumnya tidak diketahui dan berpotensi berguna.[3] Sedangkan text mining adalah suatu istilah untuk menambang suatu data yang berupa teks yang sumber datanya didapat dari dokumen. Tujuannya adalah untuk menemukan kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dicari hubungan antar dokumen yang lain.[1]

Konsep text mining biasanya digunakan untuk mengklasifikasi berbagai macam dokumen menjadi beberapa kelompok, atau mencari tingkat kemiripan atau mencari hubungan antara dokumen yang satu dengan dokumen yang lain.

2.2 Preprocessing Text

Suatu data yang berbentuk dokumen teks, mempunyai struktur kata yang tidak teratur. Sehingga dalam proses text mining diperlukan suatu proses yang dinamakan preprocessing text dengan tujuan untuk mengubah data berupa kata yang tidak terstruktur menjadi lebih terstruktur.[8] Berikut merupakan tahapan-tahapan yang dilakukan dalam preprocessing text:

1. Text Normalize
   Tahap ini dilakukan dengan mengubah teks menjadi lowercase (huruf kecil) semua. Dan dilakukan dengan penghapusan simbol dan karakter selain alphabet.
2. Tokenisasi
   Tahap ini dilakukan pemisahan kalimat ke dalam unit-unit yang lebih kecil yang dinamakan token. Tahap ini dilakukan untuk mengidentifikasi kata yang paling umum yang harus dibuang (stoplist) pada tahap stopword removal.
3. Stopword Removal
   Tahap ini dilakukan penghilangan kata-kata yang termasuk ke dalam daftar stopword. Biasanya yang terdapat pada daftar stopword yaitu kata sambung dan kata depan. Selain itu, kata dengan tingkat frekuensi kemunculannya tinggi dalam perhitungan pada kumpulan dokumen juga bisa dikatakan stopword. Pada tahap ini, penulis menggunakan library sastrawi.
4. Stemming
   Tahap ini dilakukan proses pengambilan kata dasar dari kata berimbuhan atau kata tunggal dari kata bentukan. Pada tahap ini penulis menggunakan library sastrawi untuk meningkatkan kualitas hasil pencarian pada Bahasa Indonesia.

2.3 Pembobotan Term Frequency Inverse Document Frequency (TF-IDF)

Setelah melalui tahap preprocessing text, data yang masih berupa text harus diubah ke dalam bentuk numerik. Metode pembobotan TF-IDF disini dapat digunakan untuk mengubah data yang berbentuk teks menjadi numerik dengan menentukan hubungan masing-masing kata (term) pada suatu dokumen dengan cara memberikan bobot pada masing-masing term tersebut.

TF-IDF menggabungkan dua konsep yaitu frekuensi kemunculan sebuah term di pada suatu dokumen dan inverse frekuensi dokumen yang mengandung kata tersebut.[2]

Term yang lebih sering muncul pada dokumen menjadi lebih penting karena dapat mengindikasikan topik dari dokumen. Frekuensi term *i* dalam dokumen *j* didefinisikan sebagai berikut:

$$\boldsymbol{tf_j = \frac{f_{ij}}{\max_i(f_{ij})}}$$

Persamaan 1. Frekuensi term *i* dalam dokumen *j*

Dimana fij adalah jumlah kemunculan term *i* pada dokumen *j*. Frekuensi tersebut dinormalisasi dengan frekuensi dari term yang sering muncul pada dokumen.

Inverse document frequency digunakan untuk menunjukkan *discriminative power* dari term *i*. Secara umum term yang muncul di berbagai dokumen kurang mengindikasikan untuk topik tertentu. Rumus dari inverse document frequency didefinisikan sebagai berikut:

$$\boldsymbol{idf_i = log_2\left(\frac{n}{df_i}\right)}$$

Persamaan 2. Frekuensi dokumen dari term *i*

Dimana dfi adalah frekuensi dokumen dari term *i* dan dapat diartikan juga sebagai jumlah dokumen yang mengandung term i. $Log2$ digunakan untuk meredam efek relatif terhadap $tfij$.

Weight (bobot) Wij dihitung menggunakan pengukuran TF-IDF yang didefinisikan sebagai berikut:

$$\boldsymbol{W_{ij} = tf_{ij}\ \mathrm{X}\ idf_i}$$

Persamaan 3. Perhitungan Bobot TF-IDF

Bobot paling tinggi diberikan kepada term yang sering kali muncul pada dokumen j tetapi jarang muncul dalam dokumen lain.

2.4 Cosine Similarity

Cosine similarity merupakan suatu metode yang digunakan untuk menentukan tingkat kemiripan antar teks. Pengukuran ini memungkinkan suatu dokumen untuk diberikan peringkat (ranking) sesuai dengan kemiripannya (relevansi) terhadap query.[7]

Metode pengukuran kemiripan teks yang paling populer adalah menggunakan Cosine Similarity. Metode ini mengukur nilai cosinus sudut antara dua vektor. Cosinus dari dua vector dapat diturunkan dengan menggunakan rumus *Euclidean dot product*.[6]

$$\boldsymbol{a * b\ = \|a\|\ \|b\|\ cos\,\theta}$$

Persamaan 4. *Euclidean dot product*

Diberikan dua vector dengan atribut-atribut, *A* dan *B*, nilai cosine similarity, *cos(θ)*, dinyatakan menggunakan dot product dan magnitude sebagai:

$$similarity = cos(\theta) = \frac{A + B}{\|a\| \, \|b\|} = \frac{\sum_{i=1}^{n} Ai \; x \; Bi}{\sqrt{\sum_{i=1}^{n} Ai^2} \; x \; \sqrt{\sum_{i=1}^{n} Bi^2}}$$

Persamaan 5. *Cosine Similarity*

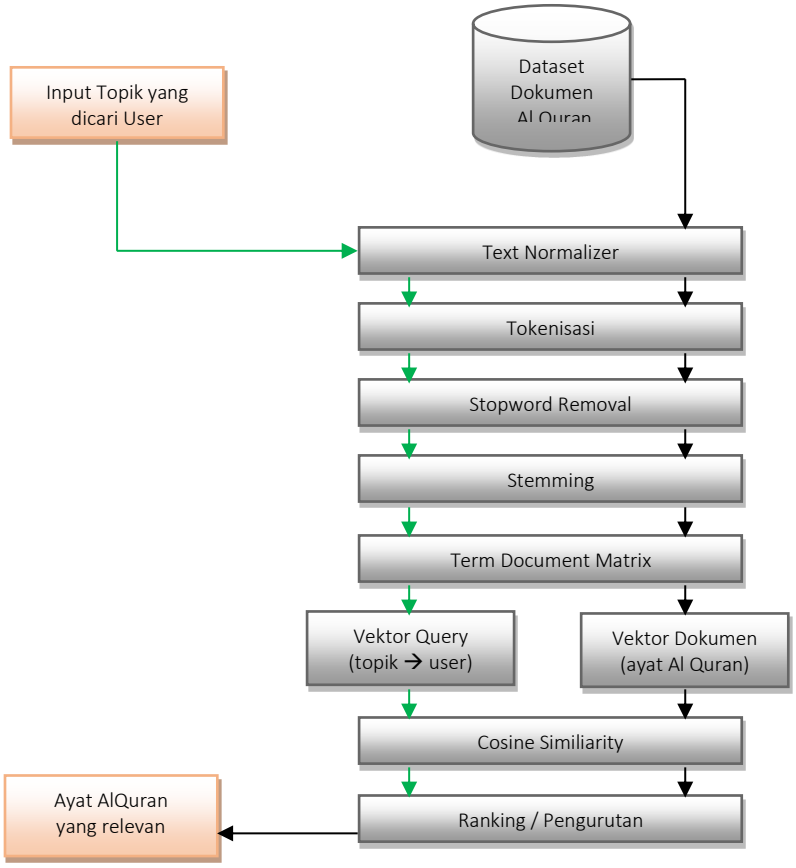
3. Metodologi Penelitian

3.1 Analisa Kebutuhan Sistem

AlQuran adalah kitab suci agama islam. Dilihat dari segi kebahasaan, "AlQuran" berasal dari bahasa Arab yang mempunyai arti "bacaan" atau "sesuatu yang dibaca berulang-ulang". Kata AlQuran merupakan bentuk kata benda dari kata kerja qara'a yang artinya membaca.[10] Penting bagi umat muslim/islam untuk bisa memahami isi yang terkandung di dalam AlQuran. Dalam penelitian ini, dibuat suatu sistem untuk menemukan ayat AlQuran berdasarkan suatu kata atau rangkaian dari beberapa kata yang akan dicari oleh pengguna. Dengan demikian, akan mempermudah umat islam khususnya, untuk menemukan topik yang dicari di dalam AlQuran

3.2 Arsitektur Sistem

Arsitektur sistem pada penelitian ini terdiri dari beberapa tahapan yang dapat dilihat pada gambar 1.



Gambar 1 Arsitektur Sistem

Pada gambar 1 dapat dilihat bahwa input yang digunakan ada dua, yaitu isi terjemahan AlQuran yang terdiri atas 30 juz atau 114 surat atau 6.236 ayat atau 152.131 kata. Dan text yang didapat dari input user berupa topik yang akan dicari.

3.3 Metode yang digunakan

Pada penelitian ini, penulis menggunakan metode pembobotan TF-IDF dan pengukurang tingkat kemiripannya menggunakan metode Cosine Similarity. Akan tetapi sebelum itu, dokumen teks terjemahan Al Quran dan input dari user diproses terlebih dahulu melalui preprocessing text dengan 4 tahapan, yaitu text normalizer, tokenisasi, stopword removal, dan stemming.

4. Hasil dan Pembahasan

Berikut akan dijelaskan terlebih dahulu hasil dari preprocessing text yang sudah dilakukan:

4.1 Text Normalize

Text normalize dimaksudkan untuk mengubah seluruh kata yang ada menjadi huruf kecil dan menghapus semua symbol dan karakter yang ada selain alphabet. Algoritma yang digunakan sebagai berikut :

Algoritma 1 Text Normalize

```
1: Ambil text ayat
2: Ganti teks menjadi huruf kecil
3: Cari dan hapus symbol selain alphabet kecuali – dan ‘
4: Hapus – yang berdiri sendiri
5: Hapus spasi lebih dari 1
```

Algoritma yang menangani text normalize menggunakan fungsi PHP `strtolower()` untuk mengubah semua teks menjadi huruf kecil. Selanjutnya menghapus semua symbol selain alphabet kecuali “-“ dan “’” serta menghapus “-“ yang berdiri sendiri dan menghapus spasi yang lebih dari 1.

"Hai anak Adam, janganlah sekali-kali kamu dapat ditipu oleh syaitan sebagaimana ia telah mengeluarkan kedua ibu bapamu dari surga, ia menanggalkan dari keduanya pakaiannya untuk memperlihatkan kepada keduanya auratnya. Sesungguhnya ia dan pengikut-pengikutnya melihat kamu dan suatu tempat yang kamu tidak bisa melihat mereka. Sesungguhnya Kami telah menjadikan syaitan-syaitan itu pemimpin-pemimpim bagi orang-orang yang tidak beriman.",

Gambar 2 Text QS Al A’raf Ayat 27 sebelum dinormalisasi

Berikut merupakan hasil normalisasinya

hai anak adam janganlah sekali-kali kamu dapat ditipu oleh syaitan sebagaimana ia telah mengeluarkan kedua ibu bapamu dari surga ia menanggalkan dari keduanya pakaiannya untuk memperlihatkan kepada keduanya auratnya sesungguhnya ia dan pengikut-pengikutnya melihat kamu dan suatu tempat yang kamu tidak bisa melihat mereka sesungguhnya kami telah menjadikan syaitan-syaitan itu pemimpin-pemimpin bagi orang-orang yang tidak beriman

Gambar 3 Text QS Al A'raf Ayat 27 setelah dinormalisasi

4.2 Tokenisasi

Tokenisasi dimaksudkan untuk memisahkan kalimat ke dalam unit-unit yang lebih kecil yang dikenal dengan istilah token. Dalam penelitian ini, proses tokenisasi dilakukan dengan menggunakan library PHP Sastrawi. Sastrawi Tokenizer merupakan library PHP agar dapat melakukan tokenization pada Bahasa Indonesia.[4] Algoritma yang dapat digunakan untuk proses tokenisasi adalah sebagai berikut:

Algoritma 2 Tokenisasi

```
1: Buat instance (object) dari class TokenizerFactory
2: Simpan pada variable tokenizer
3: Proses tokenization pada masing-masing ayat
```

Algoritma tokenisasi tersebut memanfaatkan fungsi `Tokenizer Factory()` pada library PHP Sastrawi. Hasil dari proses tokenisasi tersebut akan digunakan untuk proses selanjutnya yaitu stopword removal.

Tabel 1 Hasil Tokenisasi

No.	TOKEN
1.	hai
2.	anak
3.	adam
4.	janganlah
5.	sekali-kali
…	……
…	……
14.	mengeluarkan
15.	kedua
16.	ibu
17.	bapamu
18.	dari
19.	surga

No.	TOKEN
20.	ia
21.	menanggalkan
22.	dari
23.	keduanya
24.	pakaiannya
…	……
…	……
33.	pengikut
34.	melihat
35.	kamu
36.	dan
37.	suatu
38.	tempat

No.	TOKEN
39.	yang
40.	kamu
41.	tidak
42.	bisa
43.	melihat
…	……
…	……
51.	bagi
52.	orang-orang
53.	yang
54.	tidak
55.	beriman

4.3 Stopword Removal

Stopword removal dimaksudkan untuk memproses lebih lanjut teks ayat pada AlQuran yang sudah di tokenisasi sebelumnya. Pada tahap ini, dilakukan penghapusan pada token-token yang masuk ke dalam datfar stopword. Istilah atau kata yang masuk ke dalam daftar stopword adalah kata hubung.

Algoritma 3 Stopword Removal

```
1: Ambil daftar stopword
2: Siapkan teks ayat yang sudah ditokenisasi
3: Bandingkan teks ayat dengan daftar stopword
4: Hapus kata yang termasuk daftar stopword
5: Gabungkan teks menjadi variable string dengan pemisah spasi
```

Algoritma stopword removal diawali dengan mempersiapkan teks ayat untuk dibandingkan dengan daftar stopword. Jika terdapat kata yang sesuai, maka sistem akan menghapus kata tersebut. Hasil dari stopword removal akan digunakan untuk proses selanjutnya yaitu stemming.

Tabel 2 Hasil Stopword Removal

No.	TOKEN
1.	~~hai~~
2.	anak
3.	adam
4.	janganlah
5.	sekali-kali
…	……
…	……
14.	mengeluarkan
15.	kedua
16.	ibu
17.	bapamu
18.	~~dari~~
19.	surga

No.	TOKEN
20.	~~ia~~
21.	menanggalkan
22.	~~dari~~
23.	keduanya
24.	pakaiannya
…	……
…	……
33.	pengikut
34.	melihat
35.	~~kamu~~
36.	~~dan~~
37.	suatu
38.	tempat

No.	TOKEN
39.	~~yang~~
40.	~~kamu~~
41.	tidak
42.	~~bisa~~
43.	melihat
…	……
…	……
51.	~~bagi~~
52.	orang-orang
53.	~~yang~~
54.	tidak
55.	beriman

4.4 Stemming

Stemming dimaksudakan untuk mengubah kata berimbuhan menjadi kata dasar. Misal, *membaca* kata dasarnya *baca*, *perkebunan* kata dasarnya *kebun*, *membicarakan* kata dasarnya *bicara*, *memelihara* kata dasarnya *pelihara*. Dalam hal ini stemming dilakukan dengan menggunakan library PHP Sastrawi. Algoritma yang digunakan untuk stemming adalah sebagi berikut:

Algoritma 4 Stemming

```
1: Ambil semua ayat dari proses stopword removal
2: Lihat kata per kata pada masing-masing ayat
3: if kata = daftar kata dasar
4:    return kata
5: else
6:    Proses Stem per kata pada masing-masing ayat
7: endif
8: Simpan pada variable stemResult
```

Dalam hal ini, sistem membaca setiap kata pada masing-masing ayat. Kemudian dilakukan pencocokan pada daftar kata dasar. Apabila ditemukan maka kata tersebut

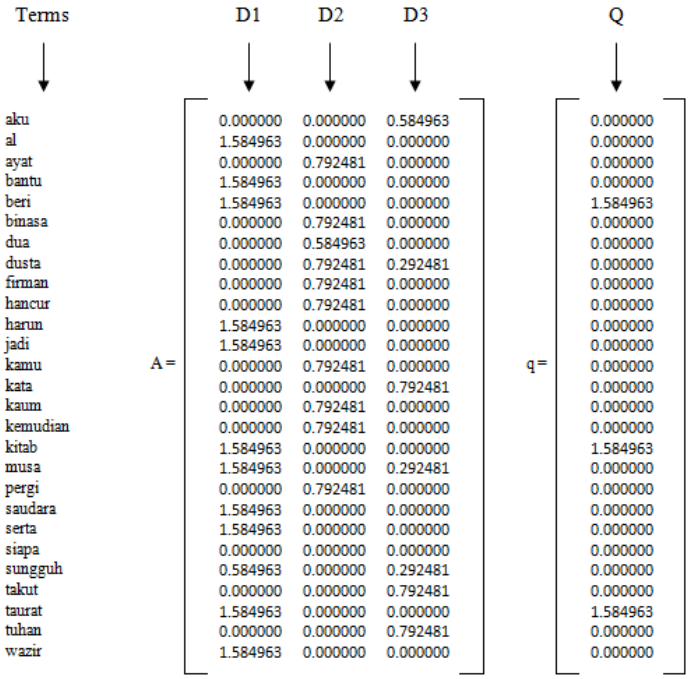
dikembalikan karena sistem menemukan bahwa kata tersebut tidak memiliki imbuhan sehingga tidak perlu dilakukan proses untuk memperoleh kata dasar yang dicari. Jika tidak ditemukan maka akan diproses menggunakan library PHP Sastrawi sehingga nantinya akan ditemukan kata dasarnya.

Tabel 3 Stemming

No.	Term	Stemming
1.	anak	anak
2.	adam	adam
3.	janganlah	jangan
…	……	……
…	……	……
14.	mengeluarkan	keluar
15.	kedua	dua
16.	ibu	ibu

No.	Term	Stemming
17.	pengikut	ikut
18.	melihat	lihat
19.	suatu	suatu
…	……	……
…	……	……
30.	sesungguhnya	sungguh
31.	menjadikan	jadi
…	……	……

Setelah preprocessing text dilanjutkan dengan pembobtan TF-IDF yang dalam hal ini dilakukan perhitungan term document matrix terlebih dahulu selanjutnya menghitung term weighting.



Gambar 4 Contoh hasil perhitungan TDM

Dilanjutkan dengan mengukur tingkat kemiripan antara topik yang dicari oleh user dengan masing-masing ayat yang terdapat di dalam AlQuran

$$sim(q.d) = \frac{q.d}{|q|.|d|}$$

$$sim(q.d1)$$

$$= \frac{(0.269156)(0.999691) + (-0.002364)(-0.001885)}{\sqrt{(0.269156)^2 + (-0.002364)^2}\sqrt{(0.999691)^2 + (-0.001885)^2}}$$

$$= 0.999974$$

$$sim(q.d2)$$

$$= \frac{(0.269156)(0.000262) + (-0.002364)(0.997868)}{\sqrt{(0.269156)^2 + (-0.002364)^2}\sqrt{(0.000262)^2 + (0.997868)^2}}$$

$$= -0.008518$$

$$sim(q.d3)$$

$$= \frac{(0.269156)(0.024875) + (-0.002364)(0.065244)}{\sqrt{(0.269156)^2 + (-0.002364)^2}\sqrt{(0.024875)^2 + (0.065244)^2}}$$

$$= 0.348018$$

Gambar 5 Contoh Hasil Perhitungan *Cosine Similarity*

Berikut merupakan hasil implementasi dari program pencarian AlQuran yang diaplikasikan dalam bentuk pemograman Web.

Gambar 6 Aplikasi Pencarian Ayat AlQuran

5. Kesimpulan

Kesimpulan yang dapat diambil dari penelitian ini adalah pemanfaatan text mining dalam pencarian ayat Al Quran dapat dilakukan dengan preprocessing text yang terdiri dari normalisasi text, tokenisasi, stopword removal dan stemming. Dilanjutkan dengan pembobotan TF-IDF dengan penghitungan term document matrix dan term weighting. Yang terakhir adalah menghitung tingkat kemiripan menggunakan Cosine Similarity. Hasilnya berupa pemograman web yang dapat digunakan oleh semua umat muslim untuk mencari topik dari sekian banyak topik yang terdapat di dalam AlQuran.

Saran yang dapat diberikan oleh penulis yaitu dengan menambahkan proses untuk pencarian sinonim dari teks yang diinputkan sehingga menghasilkan ayat AlQuran dengan lebih akurat.

Daftar Pustaka

[1] Feldman, Ronen and James Sanger. 2007. *The Text Mining Handbook.* Cambridge: Cambridge University Press. Cambridge.)

[2] Fitri, Meisya. Perancangan Sistem Temu Balik Informasi Dengan Metode Pembobotan Kombinasi Tf-Idf Untuk Pencarian Dokumen Berbahasa Indonesia. Universitas Tanjungpura : Semarang. 2013.

[3] F. Gorunescu, *Data Mining*, vol. 12. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.

[4] https://github.com/sastrawi/tokenizer

[5] J. Han and Kamber, *Data Mining : concepts and techniques. 2006.*

[6] Konchady, M. *Text Mining Application Programming*. Boston: Charles River Media.

[7] Krzysztof J. Cios, Witold Pedrycz, Roman W. Swiniarsksi, Lukasz A. Kurgan, Data Mining A Knowledge Discovery Approach, (New York, USA: Spinger, 2007)

[8] Nugroho, Eko. *Perancangan Sistem Deteksi Plagiarisme Dokumen Teks Dengan Menggunakan Algoritma Rabin-Karp*. Program studi Ilmu Komputer, Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Brawijaya. Malang. 2011.

[9] N. Ogie, Jumaidi, dan Dian Nursantika, "Perbandingan Metode Cosine Similarity dengan Metode Jaccard Similarity pada Aplikasi Pencarian Terjemahan AlQuran dalam Bahasa Indonesia" Fakultas Sains dan Teknologi, Universitas Islam Negeri Sunan Gunung Djati Bandung. 2016.

[10] [DEPAG] Departemen Agama. *Al Qur'an dan Terjemahannya.* Semarang: Toha Putra.