

Pengelompokan Tematik Al-Qur'an Menggunakan Metode *LSA* dengan Pembobotan *TF* dan *Probabilistic-IDF*

Jhody Damara, Dr. Arbi Haza Nasution

Fakultas Teknik

Program Studi Teknik Informatika

Universitas Islam Riau

E-mail : jhodydmr@student.uir.ac.id, abihazanasution@eng.uir.ac.id

ABSTRAK

Al-Qur'an adalah kitab suci umat islam yang berisi firman-firman Allah SWT, yang diwahyukan dalam bahasa Arab kepada Nabi Muhammad SAW dan membacanya bernilai ibadah. Al-Qur'an berfungsi sebagai petunjuk/pedoman bagi umat manusia dalam mencapai kebahagiaan hidup di dunia dan akhirat. Dalam mempelajari ilmu Al-Qur'an, seseorang tidak hanya diwajibkan membacanya minimal harus mengetahui tentang pengelompokan ayat-ayat berdasarkan tema yang terdapat dalam Al-Qur'an. Dimana pengelompokan tersebut dapat berfungsi untuk mengkaji lebih lanjut mengenai tema tersebut. Maka yang menjadi permasalahan adalah belum adanya korpus pengelompokan tematik Al-Qur'an yang menjelaskan tentang tema neraka yang cepat dan akurat. Penelitian ini bertujuan untuk meningkatkan efisiensi pencarian ilmiah dan mengurangi kesulitan dalam pengelompokan surah dan ayat. Dalam penelitian ini dataset yang digunakan yaitu, dataset terjemahan Al-Qur'an pada *qurandatabase.org* versi indonesia. Pengelompokan ini dilakukan menggunakan metode *Latent Semantic Analysis (LSA)* dengan pembobotan *Term Frequency* dan *Probabilistic-IDF* yang merupakan metode statistik aljabar yang mengekstrak struktur semantik yang tersembunyi dari kata dan kalimat. Penelitian ini menghasilkan pengelompokan dengan sub tema "Sifat Ahli Neraka dan Kejahatan Mereka" yang memperoleh nilai *f1-score* evaluasi Al-Qur'an Cordoba sebesar 71,85% dari penilaian matriks U, sedangkan dari pakar tematik Al-Qur'an diperoleh kontribusi penambahan 28 surah dan ayat ke dalam sub tema "Sifat Ahli Neraka dan Kejahatan Mereka" dengan peningkatan *f1-score* 9,32%.

Kata kunci : Al-Qur'an, Cordoba, *Latent Semantic Analysis(LSA)*, *Term Frequency*, *Probabilistic-Invers Document Frequency(IDF)*

1. PENDAHULUAN

Tematik Al-Qur'an adalah suatu metode untuk memahami makna tema yang terdapat dalam Al-Qur'an dengan cara menganalisis seluruh ayat Al-Qur'an tentang tema yang ingin dipahami. Contohnya pada saat penulis ingin memahami dan membahas tentang tema neraka. Pertama-tama yang penulis lakukan ialah mengumpulkan terlebih dahulu semua ayat Al-Qur'an yang membahas tentang tema neraka. Kemudian menganalisa ayat Al-Qur'an tersebut satu per-satu, bagaimana neraka menurut ayat per-ayat dalam Al-Qur'an, tujuannya untuk kemudahan seseorang

dalam memahami isi Al-Qur'an tersebut berdasarkan tema neraka secara cepat dan akurat.

Seiring dengan kemajuan ilmu pengetahuan dan teknologi, beberapa penerbit buku yang menerbitkan Al-Qur'an dan Terjemahan melampirkan jenis kategori ayat-ayat Al-Qur'an ke dalam beberapa kelas. Dalam Al-Qur'an dan Terjemahan yang diterbitkan oleh penerbit Cordoba, melampirkan bahwa dalam Al-Qur'an terdapat 15 klasifikasi ayat-ayat Al-Qur'an. Namun tidak semua ayat memiliki label kelas, ada beberapa ayat yang belum memiliki kelas.

Sehingga ayat-ayat yang belum memiliki label kelas tersebut dapat diprediksi dengan menggali informasi dari ayat yang sudah memiliki label kelasnya (Hilwah, Kudus, & Sunendiari, 2017)

Solusi yang ditawarkan oleh penulis adalah mencoba untuk membuat pengelompokan tematik Al-Qur'an tentang tema neraka berdasarkan subtema, surat, dan terjemahannya yang merujuk pada Al-Qur'an Cordoba rujukan terlengkap atau disebut juga dengan Al-Qur'an amazing, sehingga menghasilkan korpus yang nantinya bisa digunakan untuk penelitian lebih lanjut. Menurut Baker (2010:93) korpus merupakan kumpulan teks baik tulisan lisan maupun lisan yang tersimpan dalam komputer. Baker mendefinisikan korpus terdapat pada media elektronik saja. Sedangkan Menurut Setiawan (2017) korpus merupakan kumpulan tulisan yang ditulis oleh seseorang baik berupa *hard copy* dan *soft copy*. Korpus dalam bentuk *hard copy* dapat dicontohkan seperti buku, majalah, kamus, dan koran. Contoh *soft copy* dapat berupa aplikasi, *website*, kamus online, dan lain sebagainya. Jadi kesimpulannya korpus merupakan kumpulan teks baik secara lisan maupun tulisan yang ada di media cetak maupun elektronik dan dapat dijadikan sumber data.

Berdasarkan uraian di atas, maka yang menjadi permasalahan adalah belum adanya korpus pengelompokan tematik Al-Qur'an yang menjelaskan tentang tema neraka yang cepat dan akurat. Maka dari itu dibutuhkan teknologi untuk mempermudah seseorang belajar dan memahami kandungan ayat Al-Qur'an. Atas dasar latar belakang tersebut penulis tertarik untuk melakukan penelitian dengan Judul : **"Pengelompokan Tematik Al-Qur'an Menggunakan Metode LSA dengan Pembobotan TF dan Probabilistic-IDF"**.

2. LANDASAN TEORI

2.1 STUDI KEPUSTAKAAN

Penelitian pertama yang menjadi rujukan yaitu penelitian yang dilakukan oleh Husni Teja Sukmana, (2018) mengenai *"Development of Semantic Web of Indonesian Qur'an"*. Penelitian ini bertujuan untuk menyediakan sebuah korpus Al-Qur'an berbahasa Indonesia dan melakukan analisis semantik terhadap korpus yang sudah dibentuk. Korpus ini diharapkan nantinya

dapat dijadikan sebagai basis pengetahuan dalam membangun berbagai aplikasi di bidang penelusuran informasi, pengolahan bahasa manusia, dan komputasi linguistik pada domain Al-Qur'an berbahasa Indonesia.

Penelitian selanjutnya dilakukan oleh Mochamad Irfan Dary, (2015) membahas tentang *"Analisis dan Implementasi Short Text Similarity dengan Metode Latent Semantic Analysis untuk Mengetahui Kesamaan Ayat Al-Qur'an"*. Tujuan dari penelitian ini untuk membantu memahami sebuah konsep dengan cara mencari hubungan antara konsep tersebut dengan konsep lainnya menggunakan metode *Latent Semantic Analysis*. Dari hasil pengujian didapat bahwa dari berbagai dimensi yang digunakan, akurasi maksimum adalah 71% dan F-Measure 40%. Lebih baik dibandingkan *term document matrix* biasa (tanpa LSA) dengan akurasi 85% tetapi F-Measure 0%.

Penelitian terakhir dilakukan oleh Luthfiarta, Zeniarja, & Salam, (2013) membahas tentang *"Algoritma Latent Semantic Analysis (LSA) pada Peringkasan Dokumen Otomatis untuk Proses Clustering Dokumen"*. Dalam penelitian ini algoritma LSA dapat melakukan proses reduksi kalimat dengan lebih baik dibandingkan algoritma *feature based* sehingga mendapatkan hasil akurasi proses *clustering* dokumen yang lebih akurat. Hasil penelitian menunjukkan tingkat akurasi menggunakan peringkasan dokumen otomatis dengan LSA dalam proses *clustering* dokumen mencapai 71,04% yang diperoleh pada tingkat peringkasan dokumen otomatis dengan LSA 40% dibandingkan dengan hasil *clustering* tanpa peringkasan dokumen otomatis yang hanya mencapai tingkat akurasi 65,97%.

2.2 Dasar Teori

2.2.1 Tematik

Tafsir Maudhu'i (tematik) adalah mengumpulkan ayat-ayat Al-Qur'an yang mempunyai tujuan yang satu, bersama-sama membahas topic, judul, tema tertentu dengan menertibkannya sedapat mungkin sesuai dengan masa turunnya selaras dengan sebab-sebab turunnya. Kemudian memperhatikan ayat-ayat tersebut dengan penjelasan-penjelasan, keterangan-keterangan dan hubungannya dengan ayat lain kemudian mengistimbatkan hukum-hukum (Yamani, 2015).

2.2.3 Neraka

Neraka adalah tempat di alam akhirat yang diciptakan oleh Allah SWT untuk manusia dan jin yang tidak taat kepada Allah Swt dan bagi mereka yang tidak percaya bahwa Allah adalah tuhan semesta alam ketika manusia hidup di dunia. Didalam neraka penuh dengan balasan kesedihan dan penderitaan bahkan lebih lagi penghuni menjadi makhluk hina sehinah- hinanya.

2.2.4 Subtema Neraka

Berdasarkan rujukan dari Al-Qur'an Cordoba amzing yang menjadi subtema pada penelitian ini adalah sebagai berikut :

1. Nama-Nama Neraka
2. Keabadian Neraka
3. Memasuki Neraka
4. Penjaga Neraka
5. Sifat Neraka
6. Sifat Ahli Kejahatan Neraka dan Mereka
7. Siksaan Ahli Neraka Dilipatgandakan
8. Mereka yang Kekal Dalam Neraka
9. Makanan dan Minuman Ahli Neraka
10. Pohon Zaquum
11. Percakapan Ahli Neraka

2.2.5 Python

Python adalah salah satu bahasa pemrograman tingkat tinggi yang bersifat interpreter, interactive, object-oriented, dan dapat beroperasi hampir di semua platform: Mac, Linux, dan Windows. Python termasuk bahasa pemrograman yang mudah dipelajari karena sintaks yang jelas, dapat dikombinasikan dengan penggunaan modul-modul siap pakai, dan struktur data tingkat tinggi yang efisien (Prasetya & Nurviyanto, 2012).

2.2.6 Text Preprocessing

Text preprocessing merupakan suatu proses menormalisasikan teks pada dokumen sehingga informasi yang dimuat merupakan bagian yang padat dan ringkas namun tetap merepresentasikan informasi yang termuat didalamnya. Pada umumnya, proses ini dilakukan dengan cara mengeliminasi data yang tidak sesuai atau mengubah data menjadi bentuk yang lebih mudah diproses oleh sistem (Mujilahwati, 2016).

Dalam tahap ini, terdapat beberapa proses diantaranya sebagai berikut :

a. Case folding

Proses *case folding* dilakukan untuk mengubah kapitalisasi karakter (huruf) menjadi kecil untuk semua kata atau huruf mulai dari 'a' sampai 'z'. Contohnya "Saya Suka Belajar Bahasa Python" maka akan berubah menjadi "saya suka belajar bahasa python".

b. Tokenization

Proses *tokenization* dilakukan untuk pemotongan string input berdasarkan tiap kata yang menyusunnya dengan kata lain memisahkan kalimat menjadi tiap-tiap kata. Contohnya "saya adalah mahasiswa uir" maka akan berubah menjadi "saya", "adalah", "mahasiswa", "uir".

c. Stemming

Tahap *stemming* merupakan proses untuk mengubah semua kata yang telah dipilih pada proses *tokenization* menjadi kata yang berupa kata dasar. Contohnya "mempunyai" maka akan berubah menjadi "punya".

d. Filtering

Tahap *filtering* adalah tahap mengambil kata-kata penting dari hasil token dan kemudian membuang kata-kata yang tidak penting dengan menggunakan *stopwords* sehingga nantinya menghasilkan *bag of words*.

e. Stopwords

Stopwords merupakan kumpulan daftar kata-kata yang kemungkinan besar tidak akan memberikan pengaruh prediksi, seperti imbuhan dan kata ganti. Kata yang termasuk *stopwords* akan dibuang, karena kata tersebut tidak merepresentasikan isi dokumen walaupun sering muncul. Contohnya "yang", "dan", "di", "dari", "untuk", dan lain-lain.

2.2.7 Bag of Words

Bag of Words adalah sebuah model yang merepresentasikan objek secara global misalnya kalimat teks atau dokumen sebagai bag (multiset) kata tanpa memperdulikan tata bahasa bahkan urutan kata untuk menjaga keanekaragamannya. Dengan kata lain, BoW merupakan kumpulan kata-kata unik atau kata kunci dalam dokumen (Mardiana & Nyoto, 2015).

2.2.8 Latent Semantic Analysis

Latent Semantic Analysis (LSA) adalah suatu metode untuk mengekstrak sebuah tulisan dalam suatu dokumen dan kemudian mengaplikasikannya dalam perhitungan matematis. Penilaian dengan metode ini lebih kepada kata-kata yang ada dalam tulisan tanpa memperhatikan urutan kata dan tata bahasa dalam tulisan tersebut, sehingga suatu kalimat yang dinilai adalah berdasarkan kata-kata kunci yang ada pada kalimat tersebut (Winata, F., & Rainarli, E., 2016).

2.2.9 Term Frequency dan Inverse Document Frequency

Metode *Term Frequency* dan *Invers Document Frequency (TF-IDF)* merupakan metode yang digunakan menentukan seberapa jauh keterhubungan kata (*term*) terhadap dokumen dengan memberikan bobot setiap kata. Metode *TF-IDF* ini menggabungkan dua konsep yaitu frekuensi kemunculan sebuah kata di dalam sebuah dokumen dan inverse frekuensi dokumen yang mengandung kata tersebut (Fitri, 2013).

Berikut ini jenis-jenis formula pembobotan *Term Frequency* :

Pembobotan	TF Weigth
Binary	0,1
Raw count	ft, d
Term frequency	$ft, d / \sum_{t' \in d} ft', d$
Log normalization	$\log (1 + ft, d)$
Double normalization 0.5	$0.5 + 0.5 \cdot \frac{ft, d}{\max\{t' \in d\} ft', d}$

(Sumber : Poletini, 2004)

Keterangan :

$f(t, d)$ pilihan paling sederhana adalah dengan menggunakan hitungan mentah dari suatu istilah dalam dokumen yaitu, berapa kali istilah t muncul dalam dokumen d . Dimana t adalah kata kunci dan d adalah dokumen. Kemudian $\max\{t' \in d\}$ adalah Jumlah kata kunci yang sering muncul pada satu dokumen.

Berikut ini jenis-jenis formula pembobotan *Inverse Document Frequency* :

Pembobotan	IDF Weigth
Unary	1
inverse document frequency	$\log \frac{N}{nt} = -\log \frac{nt}{N}$
inverse document frequency smooth	$\log \left(\frac{N}{1+nt} \right)$

inverse document frequency max	$\log \left(\frac{\max\{t' \in d\} nt'}{1+nt} \right)$
Probabilistic inverse document frequency	$\log \frac{N-nt}{nt}$

(Sumber : Poletini, 2004)

Keterangan :

N adalah total jumlah dokumen dalam korpus, nt adalah jumlah kata kunci yang muncul pada banyaknya dokumen, dan $\max\{t' \in d\} nt'$ adalah jumlah kata kunci yang sering muncul pada banyaknya dokumen.

2.2.10 Singular Value Decomposition (SVD)

Singular Value Decomposition merupakan sebuah teknik yang biasa digunakan untuk mendekomposisi sebuah matriks dengan 3 komponen yaitu 2 matriks orthogonal serta 1 matriks yang berisi nilai-nilai singular (Arining, 2017).

SVD adalah komponen pemrosesan yang mengkompresi informasi yang berkaitan dalam jumlah besar ke dalam ruang yang lebih kecil. *SVD* merupakan proses awal dari *LSA* yang merepresentasikan isi kata dalam matriks dua dimensi yang berasal dari terjemahan Al-Qur'an tema nerakan, dimana kolom merepresentasikan surah dan ayat, dan baris mewakili *bag of words*.

Berikut ini formula dari *SVD* adalah sebagai berikut :

$$A \approx USV^T$$

Dimana :

A : matrix asal

U : *orthonormal eigenvector* dari AAT

S : matriks diagonal

V^T : *transpose* dari *orthogonal* matriks V

Dekomposisi ini memungkinkan dimensi matriks asal untuk dilakukan reduksi dimensi. Dengan proses reduksi dimensi terhadap perkalian matriks *SVD*, maka akan diperoleh penyederhanaan dan pembobotan dari matriks asal dengan mengambil sebagian besar dari struktur penting antara kata kunci dengan kalimatnya.

2.2.11 Evaluasi

Evaluasi adalah suatu proses yang sistematis untuk menentukan atau membuat keputusan, sampai sejauh mana tujuan program telah tercapai (Norman E. Grounload, 1985).

Berikut formula untuk perhitungan *precision*, *recall*, dan *f1-score*:

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{F1-Score} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

Keterangan :

True Positive (TP) = jumlah data *positif* yang terdeteksi benar.

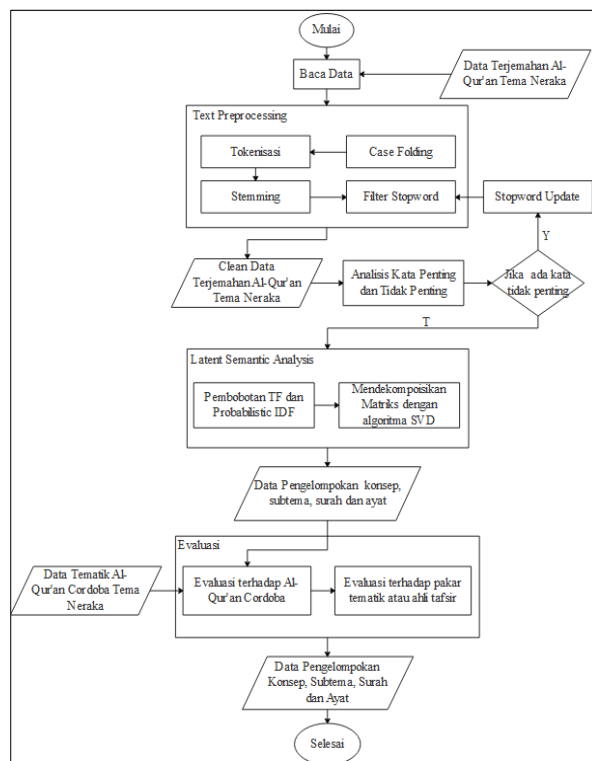
False Positive (FP) = jumlah data *negatif* namun terdeteksi sebagai data *positif*

False Negative (FN) = jumlah data *positif* namun terdeteksi sebagai data *negatif*.

3. METODE PENELITIAN

3.1 Gambaran Alur Penelitian

Alur pengelompokan Tematik Al-Qur'an dapat digambarkan melalui rancangan yang akan dibuat pada gambar dibawah ini :



Gambar 3.1 Alur Perancangan Sistem

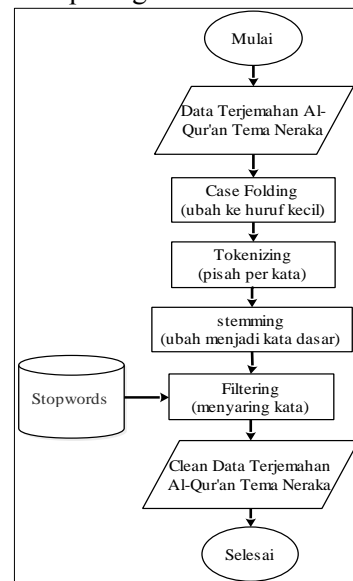
3.3 Perancangan Sistem

3.3.1 Tahapan Peprocessing

Text preprocessing merupakan tahapan dari proses awal terhadap teks untuk mempersiapkan teks menjadi data yang akan diolah lebih lanjut. Dalam penelitian ini penulis menggunakan 4 tahapan *text preprocessing* yang akan digunakan yaitu *case*

folding, *tokenization*, *stemming*, dan *filtering*.

Berikut ini tahapan *text preprocessing* dapat dilihat pada gambar 3.2 dibawah ini:

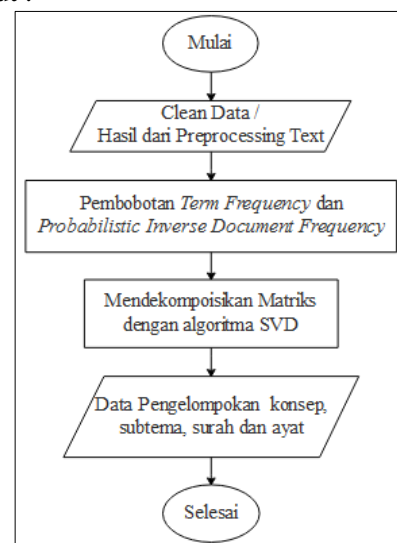


Gambar 3.1 Flowchart Text Preprocessing

3.3.3 Latent Semantic Analysis

Latent semantic analysis merupakan suatu metode untuk menemukan keterkaitan, kemiripan dan hubungan antara dokumen-dokumen, penggalan dari dokumen, dan kata-kata yang muncul pada dokumen-dokumen dengan memanfaatkan komputasi statistik untuk menggali dan merepresentasikan konteks yang digunakan sebagai sebuah arti kata untuk sejumlah korpus yang besar.

Berikut ini adalah gambaran dan penjelasan dari alur Latent Semantic Analysis yang dapat dilihat dari Gambar 3.3 sebagai berikut :



Gambar 3.3 Flowchart LSA

Pada tabel 4.4 merupakan kumpulan kata kunci atau *Bag of words* yang didapat dari proses *text preprocessing*. Pada penelitian ini

terdapat 58 kata yang menjadi *bag of words* untuk tema neraka.

4.3 Pembobotan kata terhadap Bow

4.3.1 Term Frequency

Setelah mendapatkan nilai *Bag of words* pada tiap-tiap dokumen maka tahap selanjutnya yaitu menghitung nilai *Term Frequency* dengan persamaan berikut :

$$ft, d / \sum_{t \in d} ft, d$$

Keterangan :

ft, d = banyaknya kata kunci yang muncul pada tiap-tiap dokumen

$\sum_{t \in d} ft, d$ = banyaknya kata pada tiap-tiap dokumen.

4.3.1 Invers Document Frequency

Selanjutnya melakukan proses perhitungan *Probabilistic Inverse Document Frequency* dengan persamaan berikut :

$$\log \frac{N-nt}{nt}$$

Keterangan :

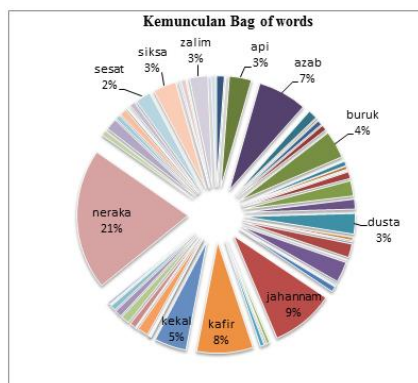
N = jumlah keseluruhan data yang terdapat pada dokumen.

nt = jumlah kemunculan kata kunci pada banyaknya dokumen.

4.3.1 TF-IDF

Setelah mendapatkan nilai *Term Frequency* dan *Probabilistic Inverse Document Frequency* terhadap *Bag of words* maka langkah selanjutnya yaitu menghitung nilai akhir dengan cara mengalikan hasil dari *Term Frequency* dengan *Probabilistic Inverse Document Frequency*.

Setelah melakukan pembobotan terhadap *term* maka dapat dilihat persentase kemunculan *bow* yang paling sering muncul dapat dilihat pada gambar 4.1 dibawah ini :



Gambar 4.1 Grafik kemunculan *bag of words*

Dari gambar 4.2 menunjukkan bahwa persentasi kata yang paling sering muncul pada tema neraka yaitu kata neraka itu sendiri sebesar 21%, kemudian disusul dengan kata jahannam 9%, kafir 8%, azab 7%, kekal 5%, dan seterusnya.

4.4 Singular Value Decomposition

Pada tahapan ini algoritma *SVD* diperlukan untuk mempermudah pengolahan data serta menyederhanakan matriks yang berukuran apa saja yang didapat setelah proses perhitungan *TF* dan *Probabilistic-IDF* dengan mendekomposisikan menjadi 3 komponen matriks yaitu menghasilkan nilai U , S , dan V^T . Tujuannya untuk menghilangkan fitur-fitur yang dirasa kurang penting dalam dokumen sehingga dapat memproduksi model data yang lebih baik.

Pada penelitian ini penulis menggunakan nilai dari matriks U dalam melakukan pengelompokan, karena nilai U merupakan proses dari terjemahan dokumen itu sendiri yang dikelompokkan berdasarkan konsep.

Berikut hasil dari nilai matriks U yang dikelompokkan oleh sistem secara otomatis dapat dilihat pada gambar 4.2 dibawah ini :

0	0.076878	-0.034896	-0.004698	0.008105	-0.020371	-0.001553	-0.002740
1	0.059712	0.044907	-0.052069	-0.042989	-0.014281	0.003551	-0.002410
2	0.009453	0.051258	0.018095	0.073979	0.003213	-0.001851	0.012971
3	0.017273	0.112143	-0.076822	-0.073111	-0.006018	-0.000977	0.002503
4	0.020750	0.134799	-0.094312	-0.089216	-0.004394	-0.007424	-0.000909
...
367	0.001461	0.003047	0.006529	-0.002394	-0.001971	0.001521	-0.005561
368	0.081225	-0.018671	0.050842	-0.027596	0.072036	-0.009663	-0.006892
369	0.146075	-0.068711	-0.014928	0.006689	-0.028352	-0.022210	0.001174
370	0.025802	0.167297	-0.070297	-0.007281	0.003094	-0.024670	-0.008093
371	0.005062	0.006844	0.000399	0.004635	-0.014440	0.190069	-0.174024
...
0	0.012481	0.015862	-0.005131	0.006994			
1	0.060658	-0.028106	0.012144	-0.008866			
2	-0.000585	0.023575	0.008186	-0.029275			
3	0.002740	0.000379	-0.003027	-0.017774			
4	-0.000891	-0.005422	-0.005656	-0.010401			
...			
367	-0.005175	-0.003688	0.003520	-0.004067			
368	-0.001885	-0.005962	-0.003474	-0.015316			
369	-0.008676	-0.014243	-0.007415	0.001045			
370	-0.003572	-0.012283	-0.004814	0.001368			
371	-0.084638	0.003690	0.093493	0.011401			

Gambar 4.2 Nilai matriks U

4.5 Pengelompokan Surah dan Ayat Per-konsep Berdasarkan Matriks U

Pada tahap ini hasil dari *singular value decomposition* nilai matriks U digunakan untuk melakukan pengelompokan surah dan ayat dari keseluruhan dokumen yang terdiri dari 11 konsep sesuai dengan subtema neraka yaitu 11 subtema. Berikut hasil pengelompokan surah dan ayat berdasarkan nilai matriks U dapat dilihat pada tabel 4.3 di bawah ini :

Tabel 4.3 Hasil pengelompokan *SVD* (U)

No	Pengelompokan	Surah dan Ayat	Jumlah
1	Konsep 0	(79, 39), (90, 20), (92, 14), (26, 96), (44, 47), (27, 90), (71, 25), (67, 7), (25, 12), (79, 36), (45, 34), (83, 16), (41, 19), (37, 163), (19, 70), (33, 66), (40, 48), (37, 23), (67, 8), (54, 48), (38, 59), (17, 63), (38, 85), (52, 13), (78, 21), (89, 23), (40, 49), (48, 13), (57, 15), (41, 24), (52, 27), (40, 46), (40, 47), (84, 12), (2, 167), (77, 32), (77, 31), (69, 31), (70, 15), (21, 99), (23, 104), (2, 119), (67, 10), (22, 51), (16, 62), (19, 72), (26, 94), (40, 7), (18, 53), (82, 14), (50, 24), (41, 40), (29, 25), (74, 27), (74, 26), (74, 42), (102, 6), (81, 12), (37, 68), (73, 12), (101, 10), (101, 9), (45, 35), (4, 145), (17, 39), (11, 106), (37, 64), (43, 77), (7, 41), (45, 10), (4, 56), (8, 14), (72, 15), (38, 61), (4, 169), (23, 103), (33, 64), (14, 29), (8, 16), (39, 60), (9, 109), (51, 13), (2, 206), (19, 86), (7, 18), (40, 60), (17, 18), (32, 13), (5, 37), (14, 30), (35, 37), (7, 38), (2, 24), (76, 4), (22, 22), (6, 70), (7, 44), (52, 18), (39, 71), (18, 102), (88, 4), (59, 17), (7, 47), (25, 11), (41, 28), (40, 43), (3, 192), (39, 72), (38, 56), (26, 91), (74, 29), (40, 11), (67, 5), (22, 72), (11, 16), (3, 12), (25, 13), (17, 8), (18, 106), (40, 6), (38, 27), (46, 20), (11, 17), (24, 57), (9, 68), (59, 3), (8, 37), (72, 23), (25, 34), (55, 43), (58, 17), (5, 10), (5, 86), (57, 19), (3, 151), (2, 126), (20, 74), (11, 119), (9, 35), (9, 17), (8, 50), (5, 72), (17, 97), (85, 10), (44, 56), (38, 57), (7, 36), (46, 34), (2, 175), (66, 6), (4, 97), (43, 74), (32, 20), (3, 116), (22, 19), (6, 128), (67, 11), (74, 31), (48, 6), (2, 39), (9, 81), (40, 76), (16, 29), (13, 5), (10, 27), (42, 45), (7, 179), (34, 42), ...	277 Surah dan Ayat

4.6 Data Evaluasi

Setelah melakukan tahap pengelompokan dengan menggunakan algoritma *SVD* berdasarkan nilai dari matriks U, langkah selanjutnya yaitu mengevaluasi hasil pengelompokan tersebut dengan data tematik Al-Qur'an Amazing atau disebut juga dengan Al-Qur'an Cordoba khusus tema neraka.

Berikut ini data tematik Al-Qur'an Cordoba tema neraka yang digunakan sebagai data evaluasi dapat dilihat pada tabel 4.4 :

Tabel 4.4 Inputan Data *Gold Standard*

No	Subtema	Surah dan Ayat	Jumlah
1	Nama-Nama Neraka	(2, 119), (2, 206), (3, 112), (3, 162), (3, 197), (4, 55), (4, 93), (4, 97), (4, 115), (4, 121), (4, 140), (4, 169), (5, 10), (5, 86), (7, 18), (7, 41), (7, 179), (8, 16), (8, 36), (8, 37), (9, 35), (9, 49), (9, 63), (9, 68), (9, 73), (9, 81), (9, 95), (9, 109), (9, 113), (11, 82), (11, 119), (13, 18), (14, 16), (14, 29), (15, 43), (16, 29), (17, 8), (17, 18), (17, 39), (17, 63), (17, 98), (18, 100), (18, 102), (18, 106), (19, 68), (19, 86), (20, 74), ...	115 Surah & Ayat

4.7 Hasil Evaluasi Cordoba

Pada tahap ini pengelompokan surah dan ayat berdasarkan nilai matriks U yang diperoleh melalui algoritma *singular value decomposition*, akan di evaluasi dengan data tematik Al-Qur'an Cordoba yang bertema neraka sesuai dengan konsep yang terbentuk yaitu sebanyak 11 konsep dengan menggunakan perhitungan *precision*, *recall*, dan *f1-score*. Berikut ini hasil evaluasi data tematik Al-Qur'an Cordoba dapat dilihat pada tabel 4.5 :

Tabel 4.5 Hasil Data Evaluasi Tematik Al-Qur'an Cordoba

Pengelompokan	No	Sub Tema	TP	FP	FN	Precision	Recall	F1-Score
Konsep 0	1	Nama-Nama Neraka	108	171	7	38%	93%	54%
	2	Keabadian Neraka	43	236	3	15%	93%	26%
	3	Memasuki Neraka	19	260	5	6%	79%	12%
	4	Penjaga Neraka	6	273	2	2%	75%	4%
	5	Sifat Neraka	57	222	32	20%	64%	30%
	6	Sifat Ahli Neraka dan Kejahatan Mereka	180	99	42	64%	81%	71%
	7	Siksaan Ahli Neraka Dilapgangkan	2	227	0	0%	100%	1%
	8	Mereka yang Kekal Dalam Neraka	20	259	5	7%	80%	13%
	9	Makanan dan Minuman Ahli Neraka	6	273	18	2%	25%	3%
	10	Pohon Zaquum	1	277	7	0%	12%	0%
	11	Percakapan Ahli Neraka	19	260	8	6%	70%	12%

Pada tabel 4.5 dapat dilihat bahwa hasil evaluasi yang dilakukan dengan menggunakan perhitungan *precision*, *recall*, dan *f1-score* masing-masing setiap konsep dan subtema memiliki nilai *f1-score* yang berbeda-beda. Dari keseluruhan data tersebut, masing-masing konsep selanjutnya di analisa kembali untuk mengambil konsep dan subtema berdasarkan hasil dari nilai *f1-score* yang terbaik atau besar dari 70% sehingga terbentuk sebuah pengelompokan yang memiliki akurasi yang lebih baik. Berikut nilai *f1-score* yang terbaik dapat dilihat pada gambar 4.4 dibawah ini :

Konsep 0 Sifat Ahli Neraka dan Kejahatan Mereka
 Surat & Ayat Tepat (TP) 180
 Surat & Ayat Tidak Tepat (FP) 99
 Surat & Ayat Tidak Tepat (FN) 42
 Precision 64.51612903225806 %
 Recall 81.08108108108108 %
 F1-Score 71.8562874251497 %

Gambar 4.4 Hasil Evaluasi *F1-Score* Terbaik

4.8 Hasil Evaluasi Pakar

Dari hasil evaluasi yang dilakukan oleh kedua pakar tersebut maka didapatkan nilai *F1-Score* yang berbeda. Perbedaan hasil evaluasi tersebut dapat dilihat pada tabel berikut :

Tabel 4.6 Perbedaan Hasil Evaluasi Kedua Pakar

Pengelompokan Konsep 0 Subtema Sifat Ahli Neraka dan Kejahatan Mereka				
Pakar	F1-Score	Cordoba	Selisih	Peningkatan
Pakar 1	88.92 %	71.85 %	17.07%	23.75%
Pakar 2	77.24 %		5.39%	7.5%

Keterangan :

Pakar 1 = Dr. Hamzah, M.Ag

Pakar 2 = Musaddad Harahap M.Pd.I

Pada tabel 4.15 dapat dilihat bahwa nilai *F1-Score* tertinggi adalah hasil evaluasi yang dilakukan oleh pakar 1 yaitu sebesar 88.92% sedangkan pakar 2 hanya 77.24%.

4.9 Perbedaan Hasil Evaluasi Al-Qur'an Cordoba dengan Hasil Evaluasi Pakar.

Berikut ini perbedaan hasil evaluasi setelah mendapatkan kontribusi dari kedua pakar dapat dilihat pada tabel dibawah ini :

Tabel 4. 72 Perbedaan Hasil Evaluasi Al-Qur'an Cordoba dengan Pakar

Pengelompokan	F1-Score		Selisih	Peningkatan
	Cordoba	Pakar		
Konsep 0 Subtema Sifat Ahli Neraka dan Kejahatan Mereka	71.85 %	78.55 %	6.7%	9.32 %

Pada tabel 4.7 dapat dilihat bahwa hasil evaluasi *f1-score* pada pengelompokan konsep 0 subtema "Sifat Ahli Neraka dan Kejahatan Mereka" meningkat setelah melakukan pengujian terhadap pakar tematik. Dari yang sebelumnya nilai *f1-score* berdasarkan Cordoba 71,85% menjadi 78,55% setelah mendapatkan kontribusi dari pakar tematik. Hasil tersebut memiliki selisih sebesar 6,7% yang di dapat dari pengurangan nilai *f1-score* dari pakar dengan nilai *f1-score* Cordoba. Selain itu hasil evaluasi tersebut juga mengalami peningkatan sebesar 9,32%. Hasil peningkatan tersebut di dapat dari angka selisih di bagi dengan nilai *f1-score* Cordoba kemudian di kali dengan 100%.

5. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan oleh peneliti melalui analisis yang telah dilakukan pada bab-bab sebelumnya maka dapat diambil kesimpulan beberapa hal sebagai berikut, yaitu :

1. Terciptanya kumpulan kata kunci atau *bag of words* tema neraka yang nantinya bisa digunakan oleh penelitian lebih lanjut.
2. Terciptanya sebuah pengelompokan tematik Al-Qur'an secara otomatis

menggunakan metode *latent semantic analysis* dengan tema neraka.

3. Pada tahap pengelompokan tematik Al-Qur'an tema neraka memperoleh hasil persentasi evaluasi menggunakan metode pengujian *Precision*, *Recall*, *F1-Score* besar dari 70%.
4. Terdapat penambahan kontribusi sebanyak 28 surah dan ayat dari hasil evaluasi pakar atau tafsir Al-Qur'an sehingga nilai *F1-Score* mendapatkan peningkatan sebesar 9,32%.

5.2 Saran

Berdasarkan hasil penelitian, agar tercapainya efektifitas dan efisiensi sesuai dengan tujuan peneliti, maka peneliti merekomendasikan atau menyarankan beberapa hal mengenai pengelompokan tematik Al-Qur'an sebagai berikut :

1. Pada penelitian ini hanya membahas tentang pengelompokan tematik Al-Qur'an tema neraka saja, penulis berharap kedepannya ada pengembang yang melakukan pengelompokan tematik Al-Qur'an untuk tema yang lainnya sehingga terbentuk pengelompokan tematik Al-Quran secara keseluruhan.
2. Untuk peneliti selanjutnya diperlukan analisis tafsir, hadist, dan asbabunnuzul sebagai pendukung untuk melakukan pengelompokan tematik Al-Qur'an.

DAFTAR PUSTAKA

- A. S., Rosa dan Shalahuddin, M. 2013. Rekayasa Perangkat Lunak Terstruktur Dan Berorientasi Objek. Informatika. Bandung.
- Al-Bahra Bin Ladjamudin. 2006. Analisis dan Desain Sistem Informasi. Graha Ilmu.Yogyakarta.
- Dary, M. I. (2015). Analisis dan Implementasi Short Text Similarity dengan Metode Latent Semantic Analysis Untuk Mengetahui Kesamaan Ayat al-Quran.Universitas Telkom.Bandung
- Fitri, M. (2013). Perancangan Sistem Temu Balik Informasi Dengan Metode Pembobotan Kombinasi *TF-IDF* Untuk Pencarian Dokumen Berbahasa Indonesia. Perancangan Sistem Temu Balik Informasi Dengan Metode

- Pembobotan Kombinasi *TF-IDF* Untuk Pencarian Dokumen Berbahasa Indonesia. Universitas Tanjungpura. Pontianak.
- Hilwah, N., Kudus, A., & Sunendiari, S. (n.d.). Klasifikasi Text Mining untuk Terjemahan Ayat-Ayat Al- Qur ' an menggunakan Metode Klasifikasi Naive Bayes. Universitas Islam Bandung.
- Jadhira, A. A., Bijaksana, M. A., & Wahyudi, B. A. (2018). Deteksi Kemiripan Bagian-bagian Terjemah Al-Qur'an dengan Menggunakan Metode Latent Semantic Analysis. Universitas Telkom. Bandung.
- Luthfiarta, A., Zeniarja, J., & Salam, A. (2013). Algoritma Latent Semantic Analysis (*LSA*) Pada Peringkat Dokumen Otomatis Untuk Proses Clustering Dokumen. Universitas Dian Nuswantoro. Semarang.
- Mardiana, T., Nyoto, R. D., Studi, P., & Informatika, T. (2015). *Kluster Bag-of-Word Menggunakan Weka*. Universitas Tanjungpura. Pontianak
- Mujilahwati, S., Studi, P., Informatika, T., Teknik, F., Lamongan, U. I., & Mining, D. (2016). PRE-PROCESSING TEXT MINING PADA DATA TWITTER. Universitas Lamongan.
- Norman E. Gronlund, (1985). *Measurement and Evaluation* : Macmillan New York.
- Nur Ichwan, Mohammad, 2004. Tafsir 'Ilmiy Memahami Al-Qur'an Melalui Pendekatan Sains Modern. Yogyakarta: Menara Kudus.
- Pangestu, A. Gelar B., Irma S. (2017). Analisis Image Watermarking Menggunakan Compressive Sensing Algoritma Orthogonal Matching Pursuit dengan Pendekatan Berbasis Discrete Cosine Transform menggunakan Singular Value Decomposition. Bandung: Universitas Telkom. Bandung.
- Poletini, N. (2004). *The Vector Space Model in Information Retrieval - Term Weighting Problem Local Term-Weighting*. Universitas of Trento.
- Prasetya, D. A., & Nurviyanto, I. (2012). Deteksi wajah metode viola jones pada opencv menggunakan pemrograman python. Universitas Muhammadiyah Surakarta.
- Siallagan, Sariadin. 2009. *Pemrograman Java*. Yogyakarta: Andi Yogyakarta.
- Sukmana, H. T. (2018). Development of Semantic Web of Indonesian Qur ' an INTERNATIONAL COLLABORATIVE RESEARCH TAHUN ANGGARAN 2015 DEVELOPMENT OF SEMANTIC WEB. Jakarta.
- Wahyuni, R. T., Prastiyanto, D., & Suprpto, E. (2017). Penerapan Algoritma Cosine Similarity dan Pembobotan *TF-IDF* pada Sistem Klasifikasi Dokumen Skripsi. Jurnal Teknik Elektro Universitas Negeri Semarang.
- Winata, F., & Rainarli, E. (2016). Implementasi Cross Method Latent Semantic Analysis untuk Meringkas Dokumen Berita Berbahasa Indonesia. Universitas Komputer Indonesia. Bandung.
- Yamani, M. T. (2015). Memahami Al-Qur'an Dengan Metode Tafsir Maudhu'i. UIN Maulana Malik Ibrahim. Malang.