

Pembobotan dan Pemeringkatan Ayat Al-Quran Berdasarkan Compound, Term Frequency dan Prinsip Pareto Untuk Membantu Hafalan

Mochamad Agung Permana¹, Eko Darwiyanto², Moch. Arif Bijaksana³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹agungpermana@students.tekomuniversity.ac.id, ²ekodarwiyanto@telkomuniversity.ac.id,

³arifbijaksana@telkomuniversity.ac.id

Abstrak

Al-Qur'an merupakan salah satu dasar hukum agama Islam yang diturunkan kepada Nabi Muhammad Shallahu 'alaihi Wasallam untuk dibaca, dipahami dan diamalkan sebagai pedoman hidup bagi umat manusia khususnya umat Islam. Untuk menghafal 114 surah yang terdiri dari 6236 bukanlah hal yang mudah. Sampai saat ini keaslian Al-Quran tetap dijaga dan terus dibukukan dengan menggunakan bahasa arab. Dibeberapa negara dengan penduduk mayoritas Muslim memiliki metode menghafal masing-masing, termasuk di Indonesia. *Term Frequency* dan *Compound Term Frequency* adalah suatu metode yang melakukan pembobotan seberapa sering sebuah *term* muncul dalam Al-Qur'an. Dengan menggunakan *Compound and Term Frequency* dapat membantu mendapatkan daftar ayat-ayat mana saja yang memiliki bobot sekitar 80%, sehingga dapat membantu proses hafalan dengan menghafal ayat-ayat tertentu saja. Hasil pembobotan untuk total 80% dengan menggunakan *Compound Term Frequency* lebih baik yaitu 46,2% dibandingkan dengan *Term Frequency* yaitu 39,2%.

Kata Kunci: *Compound Term Frequency, Term Frequency, Hafalan, Al-Qur'an*

Abstract

Al-Qur'an is one of the foundations of Islamic religious law that was revealed to the Prophet Muhammad Sallahu 'alaihi Wasallam to be read, understood and practiced as a way of life, especially Muslims. To memorize 114 chapter consisting of 6236 verses is not an easy thing. Until now, the authenticity of the Al-Quran has been preserved and is continuously recorded in Arabic. Some countries with Muslim majority populations have their own memorization methods, including in Indonesia. *Term Frequency* and *Compound Term Frequency* are methods that weight how often a term and compound words appear in a document and how often single or compound word appears in the Al-Qur'an. Using *Compound and Term Frequency* can help get a list of any verses that weigh about 80% just by memorizing certain verses, so that it can help the memorization process by memorizing certain verses. The weighting result for total of 80% using the *Compound Term Frequency* is better at 46,2%, compared to *Term Frequency* which is 39,2%.

Keywords: *Compound Term Frequency, Term Frequency, Memorizing, Al-Qur'an*

1. Pendahuluan

Aktivitas menghafal Al-Quran terus berlangsung bahkan terus semakin mapan. Pada negara yang berpenduduk mayoritas muslim sering kali dijumpai lembaga pendidikan yang berfokus untuk menghafal Al-Qur'an. Bagi umat Islam, kegiatan menghafal Al-Qur'an bukan hanya untuk memahami maknanya saja, tetapi sudah menjadi paradigma keilmuan, bahkan tujuan. Sebagian umat Islam ada yang memanfaatkan teks al-Qur'an untuk mendalami dan mengkaji secara kritis wahyu Tuhan itu. [1]

Di Indonesia, kita dapat menjumpai begitu banyak yang meminati dan memuliakan al-Qur'an. Di antara ekspresi mencintai al-Qur'an adalah dengan menghafalnya. Karena itu, kita bisa menemukan pusat pendidikan yang secara khusus mencetak anak didiknya menjadi seorang hafidz. Proses menghafal dilakukan dengan sistem *tasmi'*, yaitu mendengarkan dan memperdengarkan hafalan. Siswa mengikuti yang dibaca oleh guru, lalu menghafalkan dan pada hari selanjutnya guru tersebut melakukan evaluasi hafalan hari sebelumnya. Metode pembelajarannya tahfiz Al-Quran yang dilakukan dilakukan mengacu kepada penekanan terhadap kelancaran bacaan [2].

Seperti halnya di negara mayoritas Muslim lainnya, di Pakistan memiliki metode Panipati, siswa menghafal Al-Quran dengan mengkombinasikan beberapa tahap aktivitas yaitu *Sabak* (hafalan baru), *Nam Sabak* (mengulangi 6 Sabak), *Sabki* (ulangi 1 juz setelah 6 Sabak), *Separah* (ulangi hafalan panjang), *Mutlaah* (mengulangi dan membuat perbandingan dengan *mushaf*), Tertib Wifak (mengulang keseluruhan Al-Quran) dan *Dastar Bandi* (sertifikasi kelulusan) [3]. Di Turki, memiliki metode hafalan tersendiri, yaitu 10x *Khatm*, *Chi* (hafalan baru), *Zor Pismis*, *Kolay Pismis*, *Tekrar*, *Tekrar Hepsi Seyfa*, dan *Has*. Di Singapura, siswa melakukan hafalan seluruh ayat Al-Quran dengan cara menghafal halaman demi halaman [4]. Di Sudan dan negara-negara Afrika lain [5], menuliskan ayat-ayat Al Qur'an mungkin tidak biasa dilihat, tetapi itu merupakan praktek yang umum. Dalam *'mnemonic learning*, siswa membuat asosiasi diantara konsep yang mudah diingat dihubungkan

dengan seluruh informasi yang perlu dihafal [5]. Ada 3 jalur menghafal Al Qur'an yaitu dari juz awal sampai akhir, atau dari juz akhir hingga awal, atau dari surat-surat terpilih [6]. Menggunakan "Misri Quran" yang menggunakan font yang cantik, setiap siparah memiliki 20 halaman, dan setiap halaman berakhir dalam suatu ayat, siswa-siswa dapat menggunakannya sebagai tool *photographic memory* [7].

Dalam algoritma Bread First Search yang merupakan teori komputasi khususnya kecerdasan buatan, sesungguhnya dapat dimanfaatkan untuk mempermudah cara menghafal Al-Qu'ran. Pertama, hafalkan dulu surat-suratnya, tema utama apa yang diangkat oleh surat tersebut, cari ayat-ayat yang mendukungnya. Hafalkan dari sana, kembangkan ke ayat-ayat sebelumnya atau ayat-ayat sesudahnya.

Pada penelitian ini mencoba melakukan pendekatan lain. Dalam sebuah surah, ayat mana saja yang terpentingnya. Ahmad Hasyimi dalam [8] bahwa kaidah semantic Bahasa Arab, *ihnaab bi taukid*, pengulangan *term* menunjukkan pentingnya kata itu. *Term Frequency* dan *Compound Term Frequency* merupakan metode pembobotan dalam penambangan teks yang digunakan untuk pemberian bobot pada kata-kata tertentu terhadap sebuah dokumen, seberapa sering kata tunggal ataupun majemuk muncul di Al-Quran. Ukuran ini akan dipakai untuk menghitung seberapa penting suatu ayat, dengan cara menghitung frekuensi dari setiap katanya, kemudian menjumlahkannya. Untuk penggabungan kata majemuk menggunakan pendekatan bigram dari setiap kata dalam ayat. Jumlah *Term Frequency* dan *Compound Term Frequency* selanjutnya dapat diurutkan secara *descending*. Dari hasil pengurutan akan diketahui bobot relatif suatu ayat dengan ayat lain, termasuk ayat yang memiliki bobot terbesar, ayat yang akan dianggap terpenting di suatu surah. Bobot 80% terpilih terinspirasi dengan aturan Pareto 80/20. Vilfredo Pareto pada tahun 1896 mengamati bahwa 80% dari tanah di Italia dimiliki oleh 20% orang di Italia [9]. Dengan menghafal sedikit ayat-ayat tertentu pada Al-Quran, diharapkan dapat menghafal sebagian besar ayat-ayat penting Al-Qur'an.

1.1. Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan pada sub-bab sebelumnya, maka rumusan masalah dalam penelitian ini adalah mendapatkan ayat mana saja yang jumlah bobot *Term Frequency* dan *Compound Term Frequency* lebih dari sama dengan 80% bobot surah, serta mendapatkan kata tunggal dan kata majemuk yang berkontribusi terhadap bobot ayat yang besar. Untuk kata majemuk dilakukan dengan pendekatan bigram.

1.2. Tujuan

Tujuan pada penelitian ini adalah melakukan perangkingan dokumen dengan menggunakan metode *Term Frequency* dan *Compound Term Frequency* pada Al-Quran, lalu mendapatkan 15 kata paling sering muncul untuk dibandingkan dengan hasil riset [10], dan mendapatkan 15 kata majemuk yang sering muncul dalam Al-Quran dengan pendekatan bigram.

1.3. Organisasi Penulisan

Dalam penulisan ini, didahului dengan studi penelitian sejenis yang membahas metode penelitian sebelumnya dengan metode yang dipakai pada penelitian ini. Pada bagian selanjutnya, membahas mengenai implementasi pembobotan *Term Frequency* dan *Compound Term Frequency*. Pada bagian terakhir membahas evaluasi dan hasil nilai pembobotan pada setiap surah didalam Al-Qu'ran serta terdapat kesimpulan dan saran dari penelitian ini.

2. Studi Terkait

2.1. Text Mining

Text Mining secara luas dapat didefinisikan sebagai proses intensif dimana *user* berinteraksi dengan suatu kumpulan dokumen dari waktu ke waktu dengan menggunakan seperangkat alat analisis. Penambangan teks berupaya mengekstraksi informasi yang berguna dari sumber data melalui identifikasi dan eksplorasi dengan pola yang menarik [11]. Dokumen yang akan digunakan dalam penelitian adalah Al-Quran, perlu di *preprocessing* sebelumnya.

2.2. Term Frequency

Term Frequency adalah metode untuk menghitung bobot kata yang paling umum digunakan. Frekuensi kemunculan kata pada suatu dokumen menunjukkan seberapa penting kata itu di dalam dokumen tersebut [12]. Beberapa skema pembobotan *Term Frequency*, diantara [13]:

1. *Raw Term Frequency*
2. *Logarithmic Term Frequency*
3. *Binary Term Frequency*
4. *Augmented Term Frequency*

Dalam tabel berikut dapat merangkum skema-skema *Term Frequency*:

Tabel 1. Skema Pembobotan *Term Frequency* [13]

weighting scheme	TF weight
Binary	0, 1
Raw count	$f_{t,d}$
Log normalization	$1 + \log(f_{t,d})$
Double normalization 0.5	$0.5 + 0.5 * \frac{f_{t,d}}{\max_{\{t' \in d\}} f_{t',d}}$
Double normalization K	$K + (1 - K) \frac{f_{t,d}}{\max_{\{t' \in d\}} f_{t',d}}$

2.3. Compound Term Frequency

Compound Term disebut juga frasa, menurut Cook dalam [14], adalah satuan linguistik yang secara potensial merupakan gabungan dua kata atau lebih yang tidak mempunyai ciri-ciri klausa yaitu subjek, predikat, obyek dan keterangan. Kridalaksana dalam [14], frasa adalah gabungan dua unsur lebih yang sifatnya tidak predikatif. Menurut Hassanain dalam [14], frasa atau tarkib adalah gabungan unsur yang saling terkait dan menempati fungsi tertentu dalam suatu bentuk yang secara sintaksis sama dengan satu kata tunggal, maka gabungan unsur-unsurnya dapat diganti dengan isim atau fi'il'.

2.4. Prinsip Pareto

Prinsip Pareto atau sering disebut sebagai prinsip 80/20 dicetuskan oleh ekonom Italia Vilfredo Pareto, menyatakan bahwa untuk banyak kejadian, sekitar 80% dari efeknya disebabkan oleh 20% dari penyebabnya [9]. Dalam hal tersebut, penelitian ini mencoba menerapkan prinsip pareto untuk hafalan Al Quran. Dengan 20% dari ayat-ayat Al-Quran, siswa telah menghafal 80% dari ayat-ayat Al-Quran.

2.5. Bigram

Bigram yaitu sebuah *N-Gram* yang terdiri dari 2 item dari *sequence*. *N-Gram* adalah potongan *N*-karakter yang diambilkan dari suatu string. Blank ditambahkan pada awal dan akhir suatu string untuk mengetahui batas awal dan akhir suatu string [15]. Contohnya dalam bigram suatu string "TEXT" setelah ditambah awal dan akhir dengan "_" sebagai pengganti blank akan didapat *N-Gram* sebagai berikut: _T, TE, EX, XT, T_.

2.6. Stopword

Salah satu langkah dalam *text mining* yaitu pembuangan teks dari kata-kata yang tidak relevan dijadikan indeks. Dalam dokumen terdapat jenis kata seperti kata depan, kata sambung, kata ganti, kata sifat dan lainnya. Beberapa bagian kata tersebut tidak berpotensi dijadikan indeks dokumen karena kemunculannya tidak unik dan tidak memiliki makna. Oleh karena itu dilakukan proses penyaringan kata-kata tersebut [16]. Fadillah Z. Tala telah menyusun daftar *stoplist* yang diturunkan dari analisis kemunculan kata yang dilakukan dengan menjalankan eksperimen *korpus* Bahasa Indonesia [17].

2.7. Idhofah

Idhofah merupakan gabungan dua isim yang menyebabkan salah satu isimnya dibaca jar dikarenakan isim satunya. Anwar dalam [5] mendefinisikan idhofah sebagai pertalian antara dua perkara (dua isim) yang menyebabkan isim keduanya dibaca jar. Susunan Idhofah terdiri dari dua isim, isim yang pertama dinamakan *mudhof* dan isim yang kedua dinamakan *mudhof ilaih*. Untuk aturan dalam idhofah sendiri dibagi ke dalam bagian *mudhof* dan *mudhof ilaih* dimana masing-masing bagian ini memiliki ketentuan sendiri. Untuk *mudhof* memiliki aturan umum yaitu:

1. Tidak berawalan ۞;
2. Akhiran pada *mudhof* tidak bertanwin;
3. Membuang ۞ atau jamak pada *mudhof*.

Untuk *mudhof ilaih* memiliki aturan umum yaitu:

1. Jika diawali dengan ۞, maka berakhiran bukan tanwin;
2. Apabila tidak diawali dengan ۞, maka diakhiri dengan tanwin.

2.8. 'Athaf

Huruf '*athaf* merupakan kata penghubung yang digunakan untuk menghubungkan antara kalimat satu dengan kalimat yang lain. Anwar dalam [18] menyebutkan huruf-huruf '*athaf* ada sepuluh, yaitu:

Tabel 2. Huruf-huruf *athaf* [18]

Huruf	Arti
و	dan
أو	atau
أم	atau
ف	lalu
ثم	kemudian
حتى	sehingga
بل	tetapi
لا	bukan
لكن	tetapi
إما	atau

2.9. Huruf Jar

Huruf jar adalah salah satu kelompok huruf yang tidak bisa berdiri sendiri meskipun memiliki arti dan akan memiliki makna lengkap jika diletakkan atau dirangkaikan dengan kata benda atau isim [19].

Tabel 3. Huruf-huruf *jar* [19]

Huruf	Arti
من	dari
إلى	ke
عن	dari
على	diatas
في	didalam
رَبَّ	betapa banyak
الْبَاءُ - ب	dengan
الْكَافُ - ك	seperti
الْلامُ - ل	milik
حَتَّى	sampai
وَ	demi
تَا	demi
مُنْ dan مُدْ	sejak
خِلا	Selain/kecuali

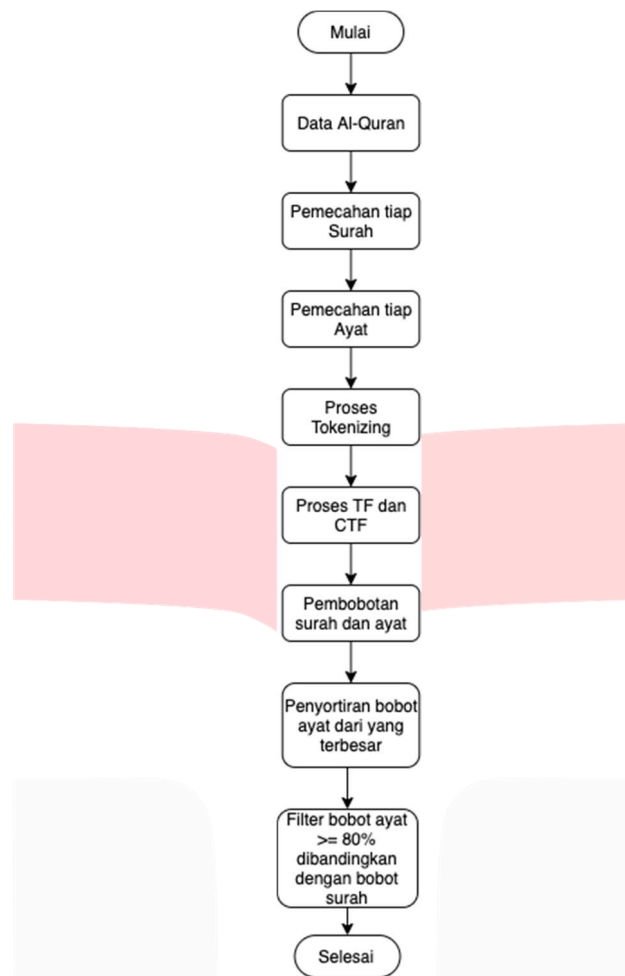
2.10. Penelitian Sebelumnya

Muazzam Ahmed Siddiqui et dalam [10] menggunakan algoritma *Latent Dirichlet Allocation* untuk menemukan topik setiap surah dalam Al-Quran dan juga *term* paling penting yang menggambarkan tema-tema. Ayat-ayat hasil penelitian ini dapat dibandingkan dengan hasil riset Muazzam.

Eko Darwiyanto dan Moch Arif Bijaksana dalam [20] dalam penelitiannya menggunakan *term frequency* pada studi kasus juz 'amma untuk memudahkan dalam menghafal Al-Quran. Dimana penelitian tersebut merupakan dasar dalam penelitian ini.

3. Rancangan Penelitian

Dalam rancangan penelitian ini dilakukan beberapa tahap dimulai dari penyiapan data dokumen yaitu Al Quran, sebagai langkah-langkah penelitian terdapat gambar 1.



Gambar 1. Flowchart tahapan penelitian

Dalam penelitian ini text Al-Quran sebagai data mentah, surah-surah pada Al-Quran dipilah, kemudian ayat-ayat juga dipilah untuk mendapatkan token-token pada setiap ayatnya. Setelah didapat token-tokennya, langkah selanjutnya yaitu mencari *frequency* setiap tokennya dalam Al-Quran. Text Al-Quran diambil dari bantuan dari situs quran.bbim.go.id. Pada proses pemisahan tiap surah beserta tiap ayatnya akan disimpan di database mysql, lalu dilakukan *tokenizing*. *Tokenizing* yaitu proses pemisahan kata-kata dari sebuah kalimat, dimana kalimat disini adalah ayat-ayat surah. Disini ada pemisahan kata-kata *compound* dengan menggunakan pendekatan bigram yang terdapat didalam tiap-tiap ayat.

Bobot tiap ayat dihitung menggunakan formula:

$$\begin{aligned} W_1(v) &= \sum TF(Token_i) \\ W_2(v) &= \sum CTF(Token_i) \end{aligned} \quad (1)$$

dimana: $W_1(v)$ adalah bobot dari ayat berdasarkan *term frequency*, $TF(Token_i)$ adalah nilai frekuensi kemunculan dari *token* ke i , i dari 1 sampai jumlah token dalam satu ayat. $W_2(v)$ adalah bobot dari ayat berdasarkan *Compound Term Frequency*, $CTF(Token_i)$ adalah *compound term frequency* dari token ke- i , i dari 1 sampai jumlah token dalam satu ayat. Jadi, pembobotan akan dihitung dalam dua versi, berdasarkan *term frequency* dan *compound term frequency* untuk setiap tokennya. Sedangkan untuk pembobotan surah menggunakan formula sebagai berikut:

$$\begin{aligned} W_1(ch) &= \sum W_1(v_i) \\ W_2(ch) &= \sum W_2(v_i) \end{aligned} \quad (2)$$

dimana: $W_1(ch)$ adalah bobot dari surah berdasarkan *term frequency*, $W_1(v_i)$ adalah bobot dari ayat ke i , i dari 1 sampai jumlah ayat dalam surah berdasarkan *term frequency*. $W_2(ch)$ adalah bobot dari surah berdasarkan perhitungan *Compound Term Frequency (CTF)*, $W_2(v_i)$ adalah bobot dari ayat ke i , i dari 1 sampai jumlah ayat dalam surah berdasarkan *compound term frequency*. Bobot surah dihitung dengan *term frequency* dan *compound term frequency* dari penjumlahan bobot setiap ayatnya.

Permasalahan muncul pada hal *stopword* dimana *stopword* dalam penambangan teks adalah kata yang dibuang dan tidak berpengaruh terhadap pembobotan *term*, tetapi teks Al-Quran harus dijaga keasliannya dan disebut kata *harf*. Disini kata *harf* tidak dihilangkan, tetapi diberi bobot 0.5 untuk menjaga keaslian teks Al-Quran.

Setelah dilakukan pembobotan menggunakan *term frequency* dan *compound term frequency*, dilakukan *sorting* jumlah bobot lebih dari atau sama dengan 80% dari bobot surahnya ayat-ayat lalu diurutkan secara *descending* dimulai dari ayat yang memiliki nilai pembobotan terbesar hingga terkecil. Setelah menemukan jumlah ayat yang mencapai total nilai bobot $\geq 80\%$ maka dilakukan formula

$$\text{Persentasi} = N(v')/N(v) \quad (3)$$

Dimana $N(v')$ adalah jumlah ayat yang memiliki nilai total pembobotan $\geq 80\%$ serta $N(v)$ yaitu jumlah keseluruhan ayat pada surah tersebut. Selanjutnya data dirangkum di dalam table summary serta nilai Persentasi dirata-ratakan untuk mengetahui hasil akhir.

4. Evaluasi

Dalam penelitian ini menggunakan pembobotan menggunakan *term frequency* menghasilkan persentasi bobot keseluruhan ayat pada surah-surah dalam seluruh Al Qur'an yang berjumlah 114 surah terdiri dari 6.236 ayat.

4.1. Hasil Pengujian

Hasil pengujian dengan metode *term frequency* yang diperoleh yaitu sebesar 39,20 % pada tabel 4 terdapat di bagian lampiran. Dalam artian untuk memahami $\geq 80\%$ seluruh isi dari seluruh Al Qur'an kita harus menghafal 39,20% dari total 6.236 ayat dengan rata-rata yang didapat untuk ayat yaitu 19. Sedangkan hasil pengujian dengan menggunakan metode *compound term frequency* diperoleh sebesar 46,20% pada tabel 5 di bagian lampiran, dalam artian untuk bisa memahami seluruh isi dari seluruh Al Qur'an harus menghafal 46,20% dari total 6.236 ayat dengan rata-rata ayat yang didapat yaitu 23.

Pengujian kedua yaitu memperhatikan seluruh ayat-ayat dalam Al Qur'an tanpa memisahkan berdasarkan setiap surahnya dengan pengurutan ayat-ayat secara *descending* mulai dari bobot ayat terbesar hingga terkecil pada metode *term frequency* menghasilkan 1.514 ayat dari total 6.236 ayat yang harus dihafalkan dan memiliki bobot $\geq 80\%$. Dengan skenario ini pada metode *term frequency* menghasilkan 24,27% dari total seluruh ayat dalam Al Qur'an yang harus dihafal. Sedangkan menggunakan *compound term frequency* menghasilkan 1.880 ayat dari total 6.236 ayat yang harus di hafal. Pada metode *compound term frequency* menghasilkan 30,14% dari total seluruh ayat Al Qur'an yang harus di hafal. Berdasarkan hasil pengujian dengan metode wawancara zoom oleh Dr Zakiy, Hafizh Quran Darul Ulul Jombang, dilampirkan di lampiran 3, beliau menyatakan bahwa hasil penelitian ini membantu untuk proses hafalan Quran, terutama kata *term* tunggal yang nilai frekuensinya tinggi, beliau menyarankan agar proses hafalan dilakukan dengan arti per katanya dengan tujuan lebih memudahkan hafalan.

4.2. Analisis Hasil Pengujian

Berdasarkan hasil pengujian yang dilakukan pada dokumen teks Al Qur'an yang berjumlah total 114 surah terdiri dari 6.236 ayat menggunakan metode pembobotan *term frequency* dan *compound term frequency* serta diuji dengan prinsip Pareto maka didapat hasil presentasi 39,20% dengan metode *term frequency*, sedangkan metode *compound term frequency* 46,20%. Dimana hasil tersebut menunjukkan menggunakan metode *term frequency* lebih baik daripada *compound term frequency*, hal ini disebabkan kata majemuk sangat berpengaruh pada nilai pembobotan pada setiap tokenya. Sementara untuk skenario kedua tanpa pembatas surah pada metode *term frequency* menghasilkan nilai 24,27% dengan total 1.514 ayat yang harus dihafal berdasarkan tabel 6 pada bagian lampiran. Berdasarkan tabel 7 pada bagian lampiran untuk skenario kedua dengan metode *compound term frequency* menyatakan 1.880 ayat yang harus dihafal dengan nilai 30,14%. Dalam artian berdasarkan pengujian 2 skenario menyatakan bahwa *term frequency* lebih baik daripada *compound term frequency*.

5. Kesimpulan dan Saran

Kesimpulan pada penelitian ini menghasilkan total ayat yang harus dihafalkan berjumlah 2.159 ayat untuk keseluruhan Al-Qur'an dengan menggunakan metode *term frequency*, 2.159 ayat ini bernilai bobot setidaknya 80% dari bobot kata-kata yang keluar di Al-Qur'an. Dibandingkan dengan menggunakan metode *compound term frequency* yang harus di hafal berjumlah 2.616, dimana dengan menggunakan metode *compound term frequency* menghasilkan jumlah ayat yang lebih banyak sebesar 457 ayat dengan pembobotan kata majemuk.

Muazzam et dalam penelitian [10] menemukan kata Allah, wala, wama dan kull paling banyak muncul di dalam Al Qur'an. Adapun 15 kata yang sering muncul dalam metode *term frequency* adalah Alloh, alladzina, inna, âmanû, al ardhî, wallâhu, rabbika, kafarû, bihi, alssamâwâtî, alnnâsu, alkitâba, walladzîna, innâ, man. Sedangkan 15 kata majemuk yang sering muncul pada metode *compound term frequency* dengan pendekatan *bigram* yaitu rabbal'alamîn, mulkussamawat, ayatilloh, ahlilkitaab, ashaabulnar, wa'udilloh, 'aalimugoib, robbulssamaawaat, amrulloh, ni'matalloh, wa'adalloh, rosululloh, 'ibadilloh, goibalssamaawaati, ni'matalloh. Untuk pengujian skenario kedua yaitu dengan mengambil seluruh ayat pada Al Qur'an tanpa membagi berdasarkan surahnya,

didapat hasil 24,27% untuk *term frequency* dengan jumlah ayat 1.514 yang perlu dihafal, sedangkan *compound term frequency* didapat hasil 30,14% dengan jumlah ayat 1.880 yang perlu dihafal. Dimana dengan menggunakan metode *term frequency* dengan hasil 24,27% mendekati dengan prinsip Pareto yaitu *Rule 80/20* sebesar 20%.

Saran untuk penelitian selanjutnya dengan metode yang sama dengan memperhatikan kata baku serta kosakata berdasarkan kamus bahasa arab sehingga hasil yang didapatkan akan lebih baik.



REFERENSI

- [1] A. Romdhoni, "Tradisi Hafalan Qur'an di Masyarakat Muslim Indonesia," *Journal of Qur'an and Hadith Studies*, vol. 1, no. 2, pp. 1-18, 2015.
- [2] M. H. Lubis, "Efektifitas Pembelajaran Tahfizhil Al-Quran Dalam Meningkatkan Hafalan Santri Di Islamic Centre Sumatera Utara," *Jurnal ANSIRU PAI*, vol. 1, no. 2, pp. 67-73, 2017.
- [3] S. Ariffin, M. Abdullah, M. and K. B. Ahmad, "Implementation Of Panipati Method On Memorization The Quran In Malaysia: A Study In Tahfiz Institute," *proceeding ITMAR*, vol. 1, pp. 406-413, 2014.
- [4] S. Ariffin, M. Abdullah and K. Ahmad, "Method on Memorization the Quran in Malaysia," *Proceedings of the International Conference on Global Business, Economics, Finance and Social Sciences*, p. Paper ID: T509, 2015.
- [5] M. I. Kenny, *Guidelines for Memorizing the Holy Quran*, Cape Town, South Africa: DPB Printers, 2009.
- [6] S. N. Khan, "A Guide to Memorizing the Qur'an," 02 10 2020. [Online]. Available: www.quranacademy.org.uk.
- [7] A. Syedna and M. M. S. Z. A. Lukmanje, "Hifz AlQuran AlMajeed, A Practical Guide," 02 10 2020. [Online]. Available: www.sfjamaat.org.
- [8] H. Ahmad, *Jawaahiru l-Balaghah fiy l-Ma'aaniy wa l-Bayaani wa l-Badii*, Beirut: Maktabah Asyriyah, 1999.
- [9] R. Koch, *The 80/20 Principle : the Secret of Achieving More with Less*, London: Nicholas Brealy Publishing, 1997.
- [10] M. A. Siddiqui, S. M. Faraz and S. A. Sattar, "Discovering the Thematic Structure of the Quran using Probabilistic Topic Model," *ICAITHQ*, pp. 234-239, 2013.
- [11] R. Feldman and J. Sanger, *The Text Mining Handbook : Advanced Approaches in Analyzing Unstructured Data*, New York: Cambridge University Press, 2007.
- [12] R. Melita, V. Amrizal, H. B. Suseno and T. Dirjam, "Penerapan Metode Term Frequency Inverse Document Frequency (TF-IDF) dan Cosine Similarity pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (STUDI KASUS: SYARAH UMDATIL AHKAM)," *Jurnal Teknik Informatika*, vol. 11, pp. 149-164, 2018.
- [13] "tf-idf," Wikipedia, [Online]. Available: <https://en.wikipedia.org/wiki/Tf-idf>. [Accessed 16 Maret 2020].
- [14] Mujianto, "Analisis Sintaksis Frasa Non Verba Dalam Buku Al Arabiyyah Lin Nasyi'in Jilid 3," 28 Maret 2012. [Online]. Available: <https://lib.unnes.ac.id/11696/1/2303408015.pdf>. [Accessed 16 Maret 2020].
- [15] B. Zaman, E. Hariyanti and E. Purwanti, "Sistem Deteksi Bahasa pada Dokumen menggunakan N-Gram," *Jurnal Multinetics*, vol. 1, no. 2, pp. 21-26, 2015.
- [16] F. Rahutomo and A. R. T. H. Ririd, "Evaluasi Daftar Stopword Bahasa Indonesia," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 6, no. 1, pp. 41-48, 2019.
- [17] F. Z. Tala, *Study of Stemming Effects on Information Retrieval in Bahasa Indonesia*, Universiteit van Amsterdam, Amsterdam, 2013.
- [18] I. T. N. Rochbani, Zukhaira and A. Miftahuddin, "Kasus Genetif (Majrurot Al-Asma) Dalam Surat Yasin (Studi Analisis Sintaksis)," *Journal of Arabic Learning and Teaching*, vol. 1, pp. 33-37, 2013.
- [19] B. Wahida, "Majrurat Al-Asma' Dalam Surah Al-Kahfi (Studi Analisis Sintaksis)," *Jurnal Al-Hikmah*, vol. 8, pp. 93-106, 2014.
- [20] E. Darwiyanto and M. A. Bijaksana, "Searching Quran Chapters Verses Weight with TF and Pareto Principle to Support Memorizing (Case Study Juz 'amma)," *ICoICT*, 2018.

- [21] F. Marir, I. Tlemsani and M. Majdalwih, "A Recursive Co-occurrence Text Mining of the Quran to Build Corpora for Islamic Banking Business Processes," *Advances in Intelligent Systems and Computing*, vol. 903, pp. 306-312, 2019.

