

Sistem Pencarian Ayat untuk Perbedaan Bunyi pada Al-Qur'an Berdasarkan Teks Kemiripan Fonetis

Agni Octavia¹, Moch. Arif Bijaksana², Kemas Muslim Lhaksana³

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

⁴Divisi Digital Service PT Telekomunikasi Indonesia

¹agnioctv@students.telkomuniversity.ac.id, ²arifbijaksana@telkomuniversity.ac.id,

³kemasmuslim@telkomuniversity.ac.id,

Abstrak

Al-Qur'an memiliki isi kandungan yang sangat banyak, sehingga sistem pencarian ayat Al-Qur'an diperlukan karena jika dilakukan secara manual akan menyulitkan. Salah satu sistem pencarian ayat Al-Qur'an yang sesuai dengan pelafalan orang Indonesia adalah Lafzi. Sistem Lafzi dapat mencari potongan ayat Al-Qur'an menggunakan kata kunci dalam aksara Latin. Lafzi telah dikembangkan menjadi Lafzi +, dimana sistem Lafzi + dapat menangani pencarian potongan ayat Al-Qur'an dengan perbedaan bunyi pada tanda berhenti. Namun, pada sistem Lafzi + belum dapat diterapkan di seluruh Al-Qur'an. Berdasarkan masalah ini, sistem perlu dikembangkan agar dapat diterapkan di seluruh Al-Qur'an. Metode yang digunakan dalam proses pencarian ayat adalah metode N-gram. N-gram yang digunakan dalam penelitian ini adalah trigram. Alur proses sistem ini pertama kali dinormalisasi dalam proses pengkodean fonetis setelah dinormalisasi kemudian tokenisasi trigram dan kemudian trigram dicocokkan antara kueri dan korpus dan dimasukkan ke dalam proses pemeringkatan untuk mendapatkan kandidat output. Hasil penelitian ini memperoleh nilai MAP (Mean Average Precision) 87% dan recall 100%.

Kata kunci : pencarian fonetis, trigram, al-qur'an

Abstract

Al-Qur'an has a lot of content, so the system of searching for verses of the Al-Qur'an is needed because if it is done manually it will be difficult. One of the search systems for the verses of the Al-Qur'an in accordance with Indonesia's pronunciation is Lafzi. The Lafzi system can search for verse fragments using keywords in Latin characters. Lafzi has been developed into Lafzi +, wherein the Lafzi + system can be used to search verses of the Al-Qur'an with different sounds on stop signs. However, the Lafzi + system cannot yet be applied throughout the Qur'an. Based on this problem, the system needs to be developed so that it can be applied throughout the Qur'an. The method used in the process of searching for the verse is the N-gram method. The N-gram used in this research is trigram. The process flow of this system is first normalized in the phonetic coding process after normalized then tokenization of trigrams and then trigrams are matched between the query and the corpus and entered into the ranking process to get an output candidate. The results of this research obtained a MAP (Mean Average Precision) value of 87% and recall of 100% .

Keywords: phonetic search, trigram, al-qur'an

1. Pendahuluan

Latar Belakang

Al-Qur'an merupakan kitab suci yang dijadikan pedoman hidup bagi umat islam. Al-Qur'an terdiri dari 30 juz, 114 surah, 6.236 ayat dan 77.845 kata [4]. Dengan isi kandungan Al-Qur'an yang sangat banyak diperlukan sistem pencarian ayat pada Al-Qur'an, karena jika melakukan pencarian secara manual akan sulit dilakukan.

Penelitian tentang sistem pencarian ayat pada Al-Qur'an sudah dilakukan sejak lama seperti Tanzil, IslamiCity dan Lafzi. Tanzil (2007) salah satu sistem aplikasi yang menyediakan fasilitas pencarian ayat pada Al-Qur'an. Namun, untuk melakukan pencarian ayat pengguna harus memasukkan kata kunci berupa aksara Arab. Hal tersebut menyulitkan bagi pengguna yang tidak bisa memasukkan kata kunci dengan aksara Arab, terutama bagi umat islam di Indonesia. Untuk mengatasi kesulitan tersebut, perlu dikembangkan sistem pencarian berdasarkan kemiripan fonetis. Sistem pencarian berdasarkan kemiripan fonetis membuat padanan antara aksara Arab dan pelafalan dalam aksara Latin. IslamiCity (2001) merupakan sistem yang dapat melakukan pencarian ayat berdasarkan kemiripan fonetis. Namun, pemadanan aksara Arab-Latin yang digunakan oleh sistem merupakan Latin-Internasional,

dimana adanya sedikit perbedaan dengan Latin-Indonesia. Lafzi (2012) salah satu aplikasi yang telah mengembangkan sistem pencarian ayat pada Al-Qur'an berdasarkan kemiripan fonetis menggunakan pemadanan aksara Arab-Latin Indonesia.

Pada sistem Lafzi dapat melakukan pencarian potongan ayat pada Al-Qur'an dengan memasukkan kata kunci dalam aksara Latin yang sesuai dengan pelafalan masyarakat Indonesia. Namun pada aplikasi tersebut belum bisa mengatasi pencarian ayat untuk perubahan bunyi pada tanda berhenti secara maksimal. Beberapa kata pada Al-Qur'an yang diikuti waqaf atau berada pada akhir ayat biasanya terjadi perbedaan dalam membacanya. Contoh pada kata $\text{é}\ll\text{é}\ddot{\text{ö}}$ "Wa a'anahu" jika diwaqafkan, maka membacanya menjadi $\text{é}\ll\ll\text{é}\ddot{\text{ö}}$ "Wa a'anah". Penelitian

dikembangkan oleh Naufal Rasyad (2019) mengenai pencarian potongan ayat Al-Qur'an dengan perbedaan bunyi pada tanda berhenti (Lafzi +) dan sistem menghasilkan MAP (Mean Average Precision) sebesar 84% dan recall sebesar 100%. Namun, pada penelitian sebelumnya masih belum diimplementasikan keseluruhan Al-Qur'an.

Oleh karena itu, pada penelitian tugas akhir ini akan menyempurnakan penelitian sebelumnya mengenai sistem pencarian ayat pada Al-Qur'an untuk perbedaan bunyi berdasarkan kemiripan fonetis. Perbedaan penelitian sekarang dan sebelumnya ialah sistem pada penelitian ini akan diimplementasikan keseluruhan Al-Qur'an dan tidak hanya mengatasi pencarian ayat untuk perubahan bunyi ditanda berhenti saja tetapi belaku untuk perubahan bunyi di tengah ayat, sehingga pada sistem sebelumnya masih ada ayat yang belum bisa ditampilkan. Metode pencarian yang digunakan adalah pencarian dengan N-grams yang diterapkan pada kode fonetis. Keuntungan menggunakan N-grams secara umum dapat dilakukan pengindeksan dengan tokenisasi N-gram pada korpus untuk membangun struktur data inverted index sehingga pencarian term tertentu dapat dilakukan dengan cepat [8].

Topik dan Batasannya

Permasalahan yang diangkat pada tugas akhir ini, yaitu bagaimana mengembangkan sistem pencarian ayat pada Al-Qur'an berdasarkan kemiripan fonetis agar dapat menangani pencarian perubahan bunyi yang dapat diimplementasikan ke seluruh Al-Qur'an. Contoh kata "Fanabazuhu" ($\text{é}\ddot{\text{ö}}\text{Y}\text{J}\text{~}$) jika kata tersebut diwaqafkan maka cara membacanya menjadi "Fanabazuh" ($\text{é}\ddot{\text{ö}}\text{Y}\text{J}$). Kata tersebut seharusnya ditemukan pada surah Al-Imran ayat 138. Namun, pada sistem Lafzi dan Lafzi+ tidak ditemukan pada ayat tersebut.

Berdasarkan dari latar belakang yang telah dijelaskan, topik-topik yang akan diangkat dalam tugas akhir ini sebagai berikut :

1. Nilai Similarity

Hasil dari tugas akhir ini berupa similarity antara kueri yang diinputkan dengan korpus. Kueri inputan dikatakan similar jika sistem dapat mengeluarkan hasil berupa surah dan ayat yang sesuai dengan gold standard dan nilai similarity yang baik berupa output yang dihasilkan terurut sesuai gold standard.

2. Input dan Output

Input dari sistem yaitu kueri berupa potongan ayat yang dicari menggunakan aksara latin dan output dari sistem yaitu menampilkan nama surah dan ayat Al-Qur'an yang mengandung kueri inputan.

Tabel 1. Input dan Output kueri Normal

No	Teks Arab	Input Teks Latin	Output Kemunculan pada Al-Qur'an (Surah: Ayat)
1.	$\text{é}\ll\text{é}\ddot{\text{ö}}$	Wa anzala	[2:22, 2:213, 3:3, 3:4, 4:113, 9:26, 14:32, 20:53, 27:60, 33:26]
2.		Bisuratin	[2:23, 10:38]
3.	$\text{é}\ll\text{!}\text{@}\text{f}$	Safa'atun	[2:48, 2:123, 2:254]

Adapun batasan-batasan dalam penelitian tugas akhir ini adalah :

1. Aplikasi yang dibangun merupakan pengembangan dari sistem yang sudah ada.
2. Pembahasan pada penelitian ini yaitu tentang bagaimana input diolah untuk menghasilkan output dan tidak membahas pembangunan sistem dari awal.

Tujuan

Tabel 2. Input dan Output Perbedaan Bunyi

No	Teks Arab	Input Teks Latin	Output Kemunculan pada Al-Qur'an (Surah: Ayat)
1.	وا انزل	Wa anzal	9:26, 14:32, 20:53, 27:60, [2:22, 2:213, 3:3, 3:4, 4:113,
2.	بِسْرٰهٖ	Bisurah	[2:23, 10:38]
3.	سَفَاۡهٖ	Safa'ah	[2:48, 2:123, 2:254]

Tujuan yang dicapai dalam penelitian tugas akhir ini adalah dapat mengembangkan sistem pencarian ayat pada Al-Qur'an untuk perubahan bunyi berdasarkan kemiripan fonetis yang dapat diimplementasikan keseluruhan Al-Qur'an dan menghitung nilai akurasi sistem yang telah dikembangkan dengan menghitung MAP dan recall.

Organisasi Tulisan

Organisasi tulisan yang ada pada jurnal adalah pendahuluan, studi terkait, sistem yang dibangun, evaluasi, kesimpulan dan daftar pustaka. Pendahuluan menjelaskan perencanaan penelitian meliputi latar belakang, tujuan, rumusan dan batasan masalah. Pada studi terkait memuat literatur yang pernah dilakukan untuk mendukung penelitian. Sistem yang dibangun merupakan perancangan pemrosesan kueri. Evaluasi menjelaskan hasil dari pengujian dan analisis terhadap hasil. Kesimpulan menjelaskan hasil akhir yang diambil dari hasil pengujian dan analisis pengujian serta memuat saran untuk dijadikan dasar pengembangan penelitian selanjutnya. Daftar pustaka berisi literatur yang membantu pengerjaan penelitian.

2. Studi Terkait

2.1 Ringkasan Penelitian

Lafzi merupakan sistem pencarian ayat pada Al-Qur'an berdasarkan kemiripan fonetis yang sesuai dengan pelafalan masyarakat Indonesia, sehingga memudahkan umat islam di Indonesia untuk melakukan pencarian potongan ayat pada Al-Qur'an. Metode pencarian pada Lafzi menggunakan pencarian dengan trigram yang diterapkan pada kode fonetis.

Lafzi + merupakan sistem pengembangan dari Lafzi. Pengembangan yang dilakukan pada Lafzi + dapat mengatasi pencarian ayat pada Al-Qur'an dengan perbedaan bunyi pada tanda berhenti. Alur proses Lafzi dengan Lafzi + tidak ada yang berbeda hanya saja adanya penambahan aturan pada proses pengodean fonetis, yaitu ditambahkan aturan untuk menangani pencarian kata dan kalimat yang diikuti tanda berhenti.

Pada Tabel 3 menunjukkan perbandingan hasil dari sistem Lafzi dan Lafzi + [10]. Pengujian terhadap sistem dilakukan secara dua kali. Skenario pertama dilakukan pengujian untuk kueri inputan yang diikuti tanda berhenti sehingga adanya perubahan dalam pengucapan. Kueri inputan pada skenario pertama berupa potongan ayat yang diwaqafkan. Contoh kata (\tilde{w} , «) "Ghisyawatun" jika diwaqafkan maka cara membacanya "Ghisyawah" maka

kueri inputan untuk skenario pertama berupa "Ghisyawah". Skenario kedua dilakukan pengujian terhadap kueri bacaan asli yang sesuai dengan aksara Arabnya. Contoh kueri inputan untuk skenario kedua "Ghisyawatu". Pengujian dilakukan dua kali karena untuk perbandingan apakah dengan menerapkan aturan pada tanda berhenti dapat mengubah hasil pencarian bacaan asli. Hasil dari Lafzi + mendapatkan nilai recall dan nilai MAP sebesar 100% dan 84%, hasil ini lebih baik dari sistem Lafzi yang hanya mendapatkan nilai recall dan nilai MAP sebesar 81% dan 65%.

Tabel 3. Tabel perbandingan Lafzi dan Lafzi +

Sistem	Perbedaan Bunyi		Bacaan Aslinya	
	Recall	MAP	Recall	MAP
Lafzi+	100%	84%	85%	58%
Lafzi	81%	65%	85%	58%

2.2 Cara Mewaqaqkan Bacaan

Saat membaca Al-Qur'an dianjurkan memperindah suara dan dituntut melakukan dalam sekali tarikan nafas untuk menghasilkan bacaan yang sempurna. Arti dalam sekali tarikan nafas pembaca mampu menyelesaikan bacaannya sampai berhenti pada sebuah tanda berhenti atau berhenti pada tengah ayat namun mengulangi pada kata atau frase sebelumnya yang masih memiliki keterkaitan maknanya. Beberapa aturan cara mewaqaqkan bacaan [9] diantaranya :

1. Jika huruf terakhir berharakat sukun (mati), maka membacanya tidak ada perubahan. Contoh lafadz فَإِذَا tetap dibaca "farghab"
2. Jika huruf terakhir berharakat fathah, kasrah, dan dhammah. Maka, huruf terakhir dibaca sukun(mati).
Contoh lafadz كُلٌّ "khalaqa" dibaca menjadi كُلٌّ "khalaq"
3. Jika huruf terakhir ta' marbutah (è), baik letaknya di tengah atau di akhir kalimat. Maka, membacanya menggantikan dengan huruf ha' (è) yang dibaca sukun.
4. Jika huruf terakhir hidup (berharakat), tetapi sebelumnya didahului dengan huruf mati (sukun), maka dua huruf tersebut dibaca mati dan huruf yang terakhir dibaca dengan suara pelan. Contoh pada lafadz بِلْهَازِلٍ ditulis "bilhazli", dibaca menjadi "bilhazl"
5. Jika di akhir kalimat didahului bacaan mad asli atau mad layyin (bacaan mad yang huruf sebelumnya berharakat fathah), maka cara membacanya yaitu dengan mematikan huruf yang terletak di akhir kalimat tersebut dengan dipanjangkan sedikit. Contoh pada lafadz أَلْحَاكِيمُ ditulis "alhakiimu", dibaca menjadi "alhakiim"
6. Ketika berhenti di akhir kalimat, tetapi huruf akhirnya berharakat fathah tanwin, maka hanya harakat fathahnya saja yang dibaca sebanyak dua harakat. Contoh pada lafadz أَفْوَاجًا ditulis "afwajaan", dan dibaca menjadi "afwaajaa"
7. Jika huruf terakhir bertaşydid, maka membacanya yaitu mematikan tanpa menghilangkan fungsi tasydidnya. Contoh pada lafadz كُلُّهَا ditulis "kholaqohunna", dibaca menjadi "kholaqohunn"
8. Hamzah di akhir kata yang ditulis di atas waw (ð) dimatikan bila waqaf, dan dibaca pendek bila washal. Contoh pada lafadz يَا فَايَا ditulis "yatafayyaw", dibaca menjadi "yatafayya"

ñ @

2.3 N-Gram

N-Gram adalah potongan karakter sebanyak N yang diambil dari string yang lebih panjang [2]. Untuk membuat N-gram yang utuh di butuhkan blank di awal dan akhir string sebagai penanda, dan biasanya blank yang sering digunakan yaitu "." underscore. Contoh string "KALIMAT" akan terdiri dari N-gram:

- Bigrams : .K, KA, AL, LI, IM, MA, AT, T.
- Trigrams : .KA, KAL, ALI, LIM, IMA, AT., T..
- Quadgrams : .KAL, ALIM, LIMA, IMAT, MAT., AT., T...

Dapat disimpulkan bahwa untuk string berukuran n akan dimiliki n unigram, n+1 bigram, n+1 trigram, n+1 quadgram, dan seterusnya. Keuntungan N-Gram dalam pencocokan string dapat bersifat resistan terhadap kesalahan tekstual karena karakteristik N-Gram sebagai bagian dari suatu string, sehingga kesalahan pada sebagian string hanya akan berakibat perbedaan pada sebagian N-Gram [2].

2.4 MAP (Mean Average Precision)

MAP merupakan nilai yang didapat dengan menghitung nilai rata-rata AP average precision [5]. AP dihitung setiap dokumen relevan yang diambil dan memperhitungkan posisi dokumen relevan dalam daftar hasil. Berikut merupakan contoh perhitungan AP tertera pada Tabel. 4 [10].

Dari tabel contoh diatas, maka didapatkan skor AP yaitu $\frac{\frac{1}{2} + \frac{2}{4} + \frac{3}{5}}{3} = 0.53$. Kategori MAP yang baik bisa dilihat dari hasil AP. Jika nilai AP tinggi artinya presisi sistem baik, karena informasi relevan yang dioutputkan oleh

sistem memiliki urutan yang berdekatan dan begitu pun sebaliknya jika nilai AP rendah artinya presisi sistem buruk, karena informasi relevan yang dioutputkan tidak memiliki urutan yang berdekatan[1].

Tabel 4. Tabel contoh untuk menghitung AP

Output ke-	Benar/Salah	Presisi	Keterangan
1	Benar	1/1	Presisi pada urutan 1
2	Salah	-	Tidak presisi
3	Salah	-	Tidak presisi
4	Benar	2/4	Presisi pada urutan 4
5	Benar	3/5	Presisi pada urutan 5

2.5 Recall

Recall merupakan jumlah dokumen relevan yang diambil dibagi dengan jumlah informasi keseluruhan yang relevan [7]. Nilai maksimum untuk recall adalah 1 dan nilai minumnya 0[10]. Jika recall bernilai 1 maka sistem berhasil melakukan pencarian sesuai dengan informasi berdasarkan kueri yang ada (gold standard). Berikut rumus untuk menghitung recall 1.

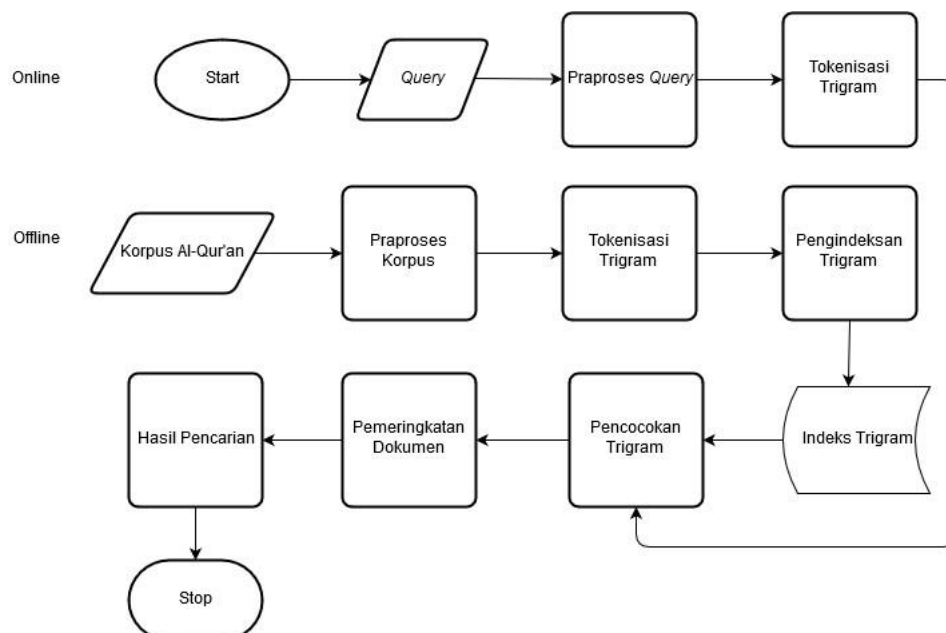
$$\text{Recall} = \frac{TP}{TP + FN} \quad (1)$$

3. Sistem yang Dibangun

Pada bagian ini akan di jelaskan mengenai hal yang terkait tentang metodologi dan desain dari sistem yang akan diimplementasikan dalam tugas akhir ini.

3.1 Gambaran Umum Sistem

Gambaran umum proses pada sistem digambarkan pada Gambar 1 Secara umum, proses pada sistem terbagi mejadi dua proses yaitu pemrosesan secara offline dan pemrosesan secara online. Pemrosesan offline dilakukan satu kali untuk membentuk indeks, sedangkan pemrosesan online dilakukan setiap dimasukkan ke sistem dan menggunakan indeks yang telah dibentuk untuk melakukan pencarian [6].



Gambar 1. Gambaran umum sistem

3.2 Praproses Query

Praproses query merupakan inputan dalam bentuk teks latin yang diubah kedalam kode fonetis agar dapat dicocokkan dengan hasil praproses pada korpus. Beberapa prosedur pengodean fonetis untuk kueri dijelaskan pada

langkah-langkah berikut[6]:

1. Substitusi Huruf Vokal

Pada aksara Arab, hanya terdapat tiga jenis huruf vokal yaitu A, I dan U [3]. Sedangkan pada aksara latin terdapat vokal lainnya yaitu E dan O. Maka dari itu huruf vokal O diganti menjadi A dan untuk huruf vokal E diganti menjadi I.

2. Penghilangan Karakter Ganda

Huruf-huruf konsonan dan vokal sama yang berdampingan dijadikan satu huruf saja.

3. Substitusi Diftong

Huruf diftong AI diubah menjadi AY, sedangkan AU diubah menjadi AW.

4. Substitusi Bacaan Idgham

Aturan pada bacaan idgham saat nun mati (نْ) bertemu dengan huruf-huruf idgham, diantaranya ya, nun, mim, wau, lam, dan ra (ي , ن , م , و , ل , dan ر). Maka huruf N dihilangkan apabila bertemu dengan huruf idgham.

5. Substitusi Bacaan Iqlab

Aturan pada bacaan iqlab saat nun mati (نْ) bertemu dengan huruf ba (ب) sehingga huruf NB akan berubah bunyi menjadi MB.

6. Substitusi Bacaan Ikhfa

Aturan pada bacaan ikhfa saat huruf nun mati (نْ) bertemu dengan huruf-huruf ikhfa. Bacaan ikhfa yaitu

bunyi huruf N yang disamarkan, dan terkadang dituliskan dengan "NG". Karakter G dihilangkan supaya setara dengan pengkodean teks Al-Qur'an.

7. Pemadanan Huruf

Pemadanan huruf dari teks latin perlu mempertimbangkan huruf arab yang direpresentasikan kedalam lebih dari satu huruf latin konsonan. Dibuatnya aturan pemadanan huruf agar dapat disesuaikan dengan pengkodean teks Al-Qur'an. Pemadanan huruf dituliskan dalam Tabel 5.

Tabel 5. Aturan Pemadanan Aksara Latin

Aksara Latin	Padanan
SH, TS, SY	S
KH, CH	H
ZH, DZ	Z
DH	D
TH	T
GH	G
NG ('ain)	X
F, V, P	F
Q, K	K
J, Z	Z
' ; (apostrof)	X

8. Penghilangan Spasi

Seluruh spasi harus dihilangkan supaya setara dengan hasil pada prosedur pengkodean teks Al-Qur'an.

9. Penambahan aturan untuk tanda berhenti

Korpus teks Al-Qur'an memiliki beberapa ciri salah satu diantaranya adalah tidak ada tanda berhenti di tengah ayat. Maka dari itu, dibuat sebuah aturan supaya bisa menemukan kata yang diikuti oleh tanda berhenti terutama untuk kata dengan akhiran ta marbuttoh (ة) yang dilafalkan menjadi ha.

Sebagai contoh, kata "ghisyawatun" (غِيْصَاوَاتُنْ) jika berada di akhir ayat atau tanda berhenti maka akan

menjadi "ghisyawah". Supaya sistem dapat menampilkan hasil untuk kueri "ghisyawah", maka dibuat aturan atau rule sebagai berikut :

- (a) Setiap kueri yang berakhiran huruf 'h' akan diubah menjadi huruf 't'. Kata "ghisyawah" akan diproses sebagai "ghisyawat"
- (b) Setiap kueri yang berakhiran huruf 'h' meskipun tidak terletak pada tanda berhenti akan tetap ditampilkan sebagai output dengan akhiran huruf 'h'. Sebagai contoh, untuk kueri "alhamdulillah" akan tetap menampilkan hasil sesuai dengan kueri tersebut.

Contoh penerapan prosedur pengkodean fonetis ini tercantum pada tabel 6 untuk kueri "maliki yaumiddiin" setelah dirubah kedalam huruf kapital. Pengkodean fonetis kueri ini dilakukan supaya menghasilkan string yang setara dengan pengkodean pada teks Al-Qur'an atau korpus.

Tabel 6. contoh pengkodean fonetis kueri

Langkah	Hasil	Deskripsi
1	MALIKI YAUMIDDIIN	tidak ada substitusi vokal E dan O
2	MALIKI YAUMIDIN	karakter ganda dihilangkan
3	MALIKI YAWMIDIN	substitusi diftong AU dengan AW
4	MALIKI YAWMIDIN	tidak ada substitusi bacaan idgham
5	MALIKI YAWMIDIN	tidak ada substitusi bacaan iqlab
6	MALIKI YAWMIDIN	tidak ada substitusi bacaan ikhfa
7	MALIKI YAWMIDIN	tidak ada pemadanan huruf
8	MALIKIYAWMIDIN	spasi dihilangkan

3.3 Tokenisasi Trigram

Proses tokenisasi trigram ialah proses pengambilan trigram dari string kode fonetis yang dihasilkan dari kueri dan korpus al-qur'an. Pembentukan trigram tidak membutuhkan penanda awal atau akhir pada string karena kueri pasti merupakan bagian dari korpus Al-qur'an [6]. Kata yang akan di tokenisasi harus diubah kedalam kode fonetis contoh kata "GHISYAWAH" diubah kedalam kode fonetis menjadi "GISAWAH". Trigram dari kata "GISAWAH" yaitu GIS, ISA, SAW, AWA, WAH.

3.4 Pencocokan Trigram

Setelah kueri ditokenisasi lalu setiap trigram dicari inverted index dari trigram korpus yang telah terindeks. Korpus yang telah terindeks terjadi pada proses offline dan tersimpan dalam database [6]. Trigram yang dihasilkan kemudian dibandingkan dengan trigram yang terdapat pada korpus kemudian jumlah trigram yang sama dihitung dan dikalikan dengan nilai ambang batas sebagai minimal skor. Jumlah kecocokan trigram lebih sama dengan minimal score akan dijadikan sebagai kandidat output.

3.5 Pemeringkatan Dokumen

Pemeringkatan dokumen secara sederhana dengan pemberian skor pada jumlah trigram yang sama antara kueri dan dokumen. Nilai maksimum ialah sebanyak jumlah trigram dan nilai minimum ialah satu karena pada proses pencarian hanya mengambil 1 trigram yang sama antara kueri dan dokumen. Kemudian pemeringkatan dokumen dengan pemberian skor pada dokumen yang posisi kemunculannya terurut dan rapat dibandingkan dengan trigram yang posisi kemunculannya teracak dan terpisah-pisah. Untuk mencari kandidat, digunakan Longest Increasing Subsequence (LIS) yang diterapkan pada posisi kemunculan trigram untuk memberi skor keterurutan [11]. Index yang memiliki keterurutan paling panjang, maka akan dijadikan kandidat hasil. LIS dari sebuah sekuens S adalah subsekuens monoton naik dari S dengan panjang maksimum. Untuk memberi skor kerapatan, dihitung rata-rata invers dari selisih antar elemen berdampingan [6] dirumuskan pada persamaan 2. Hasil dari persamaan 2 kemudian dikalikan dengan panjang LIS yang sudah didapat untuk mendapatkan skor keterurutan. Contoh, suatu bari ayat dengan index [3, 10, 2, 1, 20] maka LIS-nya [3, 10, 20] dengan panjang LIS 3 dan skor kerapatannya 0.12 maka index tersebut diberi skor $3 \times 0.12 = 0.36$

$$c = \frac{1}{n-1} \sum_{i=1}^{n-1} \frac{1}{s_{i+1} - s_i} \quad (2)$$

3.6 Hasil Pencarian

Keluaran dari sistem adalah ayat-ayat Al-Qur'an yang berisikan nomor surah, nomor ayat dan teks ayat Al-Qur'an dalam aksara Arab. Ayat yang menjadi kueri masukan akan ditambahkan highlight.

4. Evaluasi

Evaluasi dilakukan untuk mengetahui akurasi sistem berdasarkan nilai recall dan nilai MAP (Mean Average Precision).

4.1 Hasil Pengujian

Pengujian dibandingkan dengan penelitian sebelumnya dengan menambahkan 50 data test baru yang diambil dari murottal Al-Qur'an menggunakan aplikasi Tanzil. Data test dijelaskan pada analisis hasil pengujian. Pada tabel 7 menunjukkan hasil dari pengujian sistem pada penelitian ini, dimana hasilnya memiliki nilai yang lebih tinggi dari pengujian sistem sebelumnya. Hasil penelitian Lafzi + mendapatkan nilai recall sebesar 100 % dan MAP sebesar 84% [10]. Namun ketika sistem dilakukan pengujian dengan menambahkan 50 data test baru sistem mengalami penurunan nilai recall dan MAP. Penurunan nilai terjadi karena sistem Lafzi + belum bisa diimplementasikan keseluruh Al-Qur'an sehingga masih ada ayat yang belum bisa ditampilkan oleh sistem. Untuk menentukan nilai ambang batas paling optimal yang dijadikan sebagai minimal skor dalam menentukan kandidat output maka dilakukan pengujian ambang batas mulai dari 0.7 sampai 0.9. Pada Tabel 8 menunjukkan bahwa ambang batas yang paling optimal yaitu berada pada 0.7 menghasilkan MAP sebesar 76% yang memiliki hasil keterurutan paling baik dari ambang batas 0.8 dan 0.9 dan menghasilkan recall sebesar 100% dimana sistem bisa menghasilkan output sesuai gold standard.

Tabel 7. Tabel hasil pengujian

Sistem	Perbedaan Bunyi (skenario pertama)		Bacaan Aslinya (skenario kedua)	
	Recall	MAP	Recall	MAP
Lafzi++ (this work)	100%	76%	98%	77%
Lafzi+	89%	69%	92%	75%
Lafzi	76%	62%	92%	75%

Tabel 8. Tabel Nilai Ambang Batas yang Optimal

No	Nilai Ambang Batas	MAP	Recall
1	0.9	74%	97%
2	0.8	75%	100%
3	0.7	76%	100%

4.2 Analisis Hasil Pengujian

Pengujian sistem dilakukan menggunakan 100 kueri yang merupakan 50 data test penelitian sebelumnya dan 50 data test pada penelitian ini. Skenario pertama dilakukan pengujian untuk kueri inputan yang bacaannya dapat diwaqafkan atau terdapat aturan nun washal sehingga adanya perubahan dalam membacanya yang tidak sesuai dengan aksara Arabnya. Berikut contoh kueri pada skenario pertama dapat dilihat pada tabel 9. Kueri inputan "Fanabazuh" (فَنَابَازُوحٌ) merupakan salah satu contoh kueri yang dapat diwaqafkan dan kueri inputan "Masalanil

مَسَالَانِيلٌ) merupakan salah satu contoh kueri yang terdapat aturan nun washal. Dimana pada kedua contoh kueri tersebut adanya perbedaan dalam pengucapan yang tidak sesuai dengan aksara Arab. Pada kueri "Masalanilkaum" seharusnya ditemukan pada surah Al-A'raf ayat 177 tetapi pada sistem Lafzi dan Lafzi + tidak ditemukan karena adanya perbedaan trigram yang dihasilkan antara kueri inputan dan korpus sehingga menyebabkan kueri inputan tidak ditemukan oleh sistem. Dengan menambahkan fungsi pada proses pengodean fonetis dan memperbaiki proses pada pencarian index maka sistem dapat menemukan hasil pencarian dan meningkatkan nilai recall dan MAP.

Tabel 9. Contoh Kueri Pengujian Pertama

No	Teks Arab	Teks Latin	Recall		
			Lafzi ++	Lafzi +	Lafzi
1	(ٱ ٱ ٱ ٱ ٱ)	Fanabazuh	1	0	0
2	(ٱ ٱ ٱ ٱ)	Wa a'anah	1	0	0
	(ٱ ٱ)	Masalanil kaum	1	0	0

Skenario kedua dilakukan pengujian terhadap kueri bacaan asli yang sesuai dengan aksara Arabnya. Berikut contoh kueri untuk pengujian skenario kedua dapat dilihat pada tabel 10. Hasil pengujian skenario kedua didapat nilai recall 98% dan MAP 77% lebih tinggi dari sistem sebelumnya. Hal ini karena adanya penambahan fungsi pada proses pengkodean fonetis dan memperbaiki proses pada pencarian index sehingga menghasilkan hasil pencarian dimana pada sistem Lafzi dan Lafzi + tidak ditemukan dan pada sistem Lafzi ++ ditemukan.

Tabel 10. Contoh Kueri Pengujian Kedua

No	Teks Arab	Teks Latin	Recall		
			Lafzi ++	Lafzi +	Lafzi
1	(ٱ ٱ ٱ ٱ)	Nhubnahu	1	1	1
2	(ٱ ٱ ٱ)	Wa a'anahu	1	0	0
	(ٱ ٱ)	Masala alkaumu	1	0	0

5. Kesimpulan

Berdasarkan hasil pengujian yang diperoleh dan dianalisis dapat ditarik kesimpulan bahwa sistem pencarian ayat pada Al-Qur'an untuk perbedaan bunyi berdasarkan kemiripan fonetis yang telah dikembangkan tidak menurunkan nilai recall dan MAP sistem sebelumnya, dimana pada sistem sebelumnya mendapatkan nilai recall 100% dan MAP 84% meskipun saat diuji dengan 50 data test baru menurunkan nilai dan setelah dikembangkan menjadi meningkat kembali yaitu dengan nilai recall 100% dan MAP 87%. Saran untuk penelitian selanjutnya dapat menanganikan pencarian ayat berdasarkan perubahannya bunyi secara lintas ayat.

Daftar Pustaka

- [1] P. A. Arsaningtyas, M. A. Bijaksana, and S. Al Faraby. Sistem pencarian ayat al-quran berdasarkan kemiripan ucapan menggunakan algoritma soundex dan damerau-levenshtein distance. Universitas Telkom. Bandung, 2018.
- [2] W. B. Cavnar, J. M. Trenkle, et al. N-gram-based text categorization. In Proceedings of SDAIR-94, 3rd annual symposium on document analysis and information retrieval, volume 161175. Citeseer, 1994.
- [3] S. Hadi, S. C. Soeratno, M. Ramlan, and I. D. P. Wijana. Perubahan Fonologis Kata-kata Serapan dari Bahasa Arab dalam Bahasa Indonesia. Gadjah Mada University, 2003.
- [4] B. Hammo, A. Sleit, and M. El-Haj. Effectiveness of query expansion in searching the holy quran. 2007.
- [5] S. Han, D. He, Z. Yue, and J. Jiang. Contextual support for collaborative information retrieval. In Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval, pages 33–42, 2016.
- [6] M. A. Istiadi. Sistem pencarian ayat al-qur'an berbasis kemiripan fonetis, 2012.
- [7] D. Kelly. Methods for evaluating interactive information retrieval systems with users. Foundations and trends in Information Retrieval, 3(1–2):1–224, 2009.
- [8] M.-S. Kim, K.-Y. Whang, J.-G. Lee, and M.-J. Lee. n-gram/2l: A space and time efficient two-level n-gram inverted index structure. In Proceedings of the 31st international conference on Very large data bases, pages 325–336, 2005.

- [9] Nadiazhr. 13 macam tanda waqaf yang wajib kamu ketahui. Accessed: 2019-06-22.
- [10] N. Rasyad, M. A. Bijaksana, and K. M. Lhaksana. Pencarian potongan ayat al-qur'an dengan perbedaan bunyi pada tanda berhenti berdasarkan kemiripan fonetis. *Jurnal Linguistik Komputasional*, 2(2):56–61, 2019.
- [11] D. Romik. *The surprising mathematics of longest increasing subsequences*, volume 4. Cambridge University Press, 2015.