OPEN ACCESS
Indo-JC Indonesia Journal of Computing
ISSN 2460-9056
socj.telkomuniversity.ac.id/indojc

Ind. Journal on Computing
Vol. 4, Issue. 3, Dec 2019. pp. 23-36
doi:10.21108/indojc.2019.4.3.351

Pembangunan Korpus dari Rangkaian Kata yang Berulang pada Alquran

Miftahul Adnan Rasyid [1], Moch. Arif Bijaksana [2], Ibnu Asror [3]

Fakultas Informatika, Universitas Telkom
Bandung, Indonesia

[1] madnanrasyid@student.telkomuniversity.ac.id

[2] arifbijaksana@telkomuniversity.ac.id

[3] iasror@telkomuniversity.ac.id

Abstract

One way to understand the Quran is to do interpretations correctly and not deviate, is by paying attention to the word editor that used in composing verses of the Quran. In this way, another verse can be found that has a similar set of words. One way to search for the same verse based on these words is to use the Longest Common Subsequence (LCS) approach which can find the longest shared word sequence of a text pair. The results of the same verse search are then collected to become a corpus which is expected to help humanity in interpreting the Quran. This research produces a system that can search for the same verse using the LCS approach, then results will be collected into a corpus based on the LCS results. The average results obtained from several tests that have been done are Arabic data getting the precision value is 46.84%, the recall value is 96.13%, and the f1-score value is 62.96%. While Indonesian Data getting the precision value is 40.57%, the recall value is 97.56%, and the f1-score value is 57.04%.

Keywords: Corpus, Longest Common Subsequence, The Quran, Word Sequence.

Abstrak

Salah satu cara untuk memahami Alquran adalah dengan melakukan penafsiran yang benar dan tidak menyimpang, yaitu dengan memperhatikan redaksi kata yang digunakan dalam merangkai ayat-ayat Alquran. Dengan cara tersebut, maka dapat ditemukan ayat lainnya yang memiliki rangkaian kata yang menyerupai. Salah satu cara untuk mencari ayat yang sama berdasarkan rangkaian kata tersebut adalah dengan menggunakan pendekatan *Longest Common Subsequence* (LCS) yang dapat mencari rangkaian kata terpanjang bersama dari suatu pasangan teks. Hasil dari pencarian ayat yang sama ini kemudian dikumpulkan hingga menjadi korpus yang diharapkan dapat membantu umat manusia dalam menafsirkan Alquran. Penelitian ini menghasilkan suatu sistem yang dapat mencari ayat yang sama menggunakan pendekatan LCS, kemudian hasilnya akan dikumpulkan menjadi suatu korpus berdasarkan hasil LCS. Hasil rata-rata yang diperoleh dari beberapa pengujian yang telah dilakukan adalah Data Arab mendapatkan nilai *precision* adalah 46.84%, nilai *recall* adalah 96.13%, dan nilai *f1-score* adalah 62.96%. Sedangkan untuk Data Indonesia mendapatkan nilai *precision* adalah 40.57%, nilai *recall* adalah 97.56%, dan nilai *f1-score* adalah 57.04%.

Kata Kunci: Alquran, Korpus, *Longest Common Subsequence*, Rangkaian Kata.

I. Pendahuluan

ALQURAN adalah suatu kitab mukjizat melalui ungkapannya yang demikian indah memukau, terdiri dari huruf-huruf seperti *alif- lam- ra* dan berfungsi sebagai kitab suci yang diturunkan atas izin

Received on Nov 2019. Accepted on Dec 2019

Allah SWT kepada Nabi Muhammad SAW untuk menjadi pedoman hidup bagi umat manusia yang mengeluarkan mereka dari gelapnya kekufuran menuju cahaya keislaman yang terang benderang [15] [10]. Untuk menjadikannya sebagai pedoman hidup, Alquran harus dipahami dengan melakukan penafsiran secara benar dan tidak menyimpang. Salah satu caranya adalah dengan memperhatikan redaksi kata, yaitu struktur kata yang digunakan dalam merangkai atau menyusun ayat-ayat pada Alquran [5]. Dengan melihat redaksi atau rangkaian kata tersebut maka dapat ditemukan ayat lainnya yang memiliki rangkaian kata yang menyerupai. Contohnya seperti pada Quran Surah *al-Fatihah* ayat 1 (QS:1:1) yang artinya *Dengan menyebut nama Allah yang Maha Pemurah lagi Maha Penyayang* dengan Quran Surah *al-Fatihah* ayat 3 (QS:1:3) yang artinya *Maha Pemurah lagi Maha Penyayang*.

Untuk membangun suatu korpus dari rangkaian kata yang berulang pada Alquran ini haruslah dilakukan pencarian ayat yang memiliki struktur kata yang sama pada ayat lainnya dan selanjutnya hasil tersebut dikumpulkan. Pencarian ayat pada Alquran telah banyak dilakukan oleh beberapa penelitian sebelumnya dengan menggunakan metode ataupun pendekatan yang beragam, dan memiliki kesamaan dengan penelitian kali ini [2] [9] [12]. Hasil yang diperoleh dari beberapa penelitian tersebut cukup baik namun hanya menyajikan hasil pencarian ayat berupa tema ayat, nomor ayat, dan nama surah berdasarkan masukkan dari pengguna. Berbeda dari penelitian sebelumnya, penelitian kali ini adalah melakukan pencarian ayat yang sama dengan ayat lainnya sekaligus menghimpun atau mengelompokkan hasil pencarian tersebut menjadi sebuah korpus yang berisi himpunan ayat yang sama berdasarkan rangkaian kata terpanjang yang digunakan dalam menyusun ayat tersebut.

Berdasarkan permasalahan di atas, dibuatlah suatu sistem yang dapat mencari ayat yang memiliki redaksi kata yang sama dengan ayat lainnya menggunakan pendekatan *Longest Common Subsequence* (LCS), yaitu suatu metode untuk mencari rangkaian data terpanjang bersama dari suatu kumpulan data [7] [6], dalam hal ini data tersebut adalah ayat Alquran. Data yang digunakan pada penelitian ini adalah seluruh ayat pada Alquran dan terjemahan Alquran berbahasa Indonesia yang terlebih dahulu dilakukan *Text Preprocessing* berupa penggunaan bentuk teks *lowercase*, penghapusan *stopword*, dan penghapusan semua karakter yang tidak dibutuhkan. Dari hasil tersebut selanjutnya dikumpulkan menjadi suatu korpus berupa himpunan ayat berdasarkan rangkaian kata terpanjang bersama. Data akhir tersebut dievaluasi dengan cara dibandingkan dengan data *gold-standard* yang dibuat oleh penulis dengan ketentuan bahwa ayat yang diprediksi sama dengan ayat lainnya adalah ayat yang memiliki rangkaian kata terpanjang dengan jumlah rangkaian kata lebih dari atau sama dengan selisih antara jumlah kata pada ayat pembanding dan jumlah rangkaian kata terpanjang tersebut. Hal tersebut dikarenakan penulis berasumsi bahwa jumlah rangkaian kata terpanjang yang baik adalah yang jumlah katanya sama dengan atau lebih dominan dibandingkan sisa kata dari teks tersebut. Evaluasi yang dilakukan adalah dengan menghitung *precision*, *recall*, dan *f1-score*. Korpus ini diharapkan dapat membantu umat manusia dalam menafsirkan Alquran dengan cara yang benar dan tidak menyimpang.

II. Studi Terkait

Pembangunan korpus dari rangkaian kata yang berulang pada Alquran ini diawali dengan melakukan pencarian ayat yang sama dengan memperhatikan redaksi kata yang digunakan dalam menyusun atau merangkai ayat tersebut. Pencarian ayat tersebut telah dilakukan oleh beberapa penelitian dan memiliki keterkaitan dengan penelitian yang sedang dilakukan.

Pada tahun 2012, telah dilakukan suatu penelitian tentang pencarian ayat Alquran berdasarkan kata atau frase oleh Agus Sofiyan, et al. dengan menggunakan teknik *Inexact String Matching* [2]. Teknik tersebut adalah teknik pencarian ayat yang dianggap sesuai jika pengguna memasukkan kata atau frase yang akan dicari dengan benar. Penelitian ini memadukan teknik *stemming* yang berperan sebagai *Preprocessing* berupa penghapusan semua imbuhan baik prefiks maupun sufiks, dan teknik *Exact String Matching* yang dapat mencocokkan *string* secara tepat dengan susunan karakter dalam *string* yang dicocokkan memiliki jumlah maupun urutan karakter yang sama.

Pada tahun yang sama, Muhammad Abrar Istiadi telah melakukan penelitian tentang pencarian ayat Alquran dengan menggunakan cara yang berbeda, yaitu dengan menggunakan pelafalan kata dalam aksara Latin [12]. Penelitian ini bertujuan untuk membangun sistem pencarian ayat Alquran berbasis kemiripan

fonetis berkenaan dengan aturan bacaan Alquran (tajwid) untuk mencocokkan antara teks-teks Alquran dalam aksara Arab dengan representasi pelafalan orang Indonesia. Metode pencarian ayat yang digunakan pada penelitian ini adalah metode *n-gram* yang digabungkan dengan pengodean fonetis. Menurut Aqeel, et al. [4], nilai n atau jumlah *gram* yang sering digunakan adalah 2 dan 3 (*bigram* dan *trigram*).

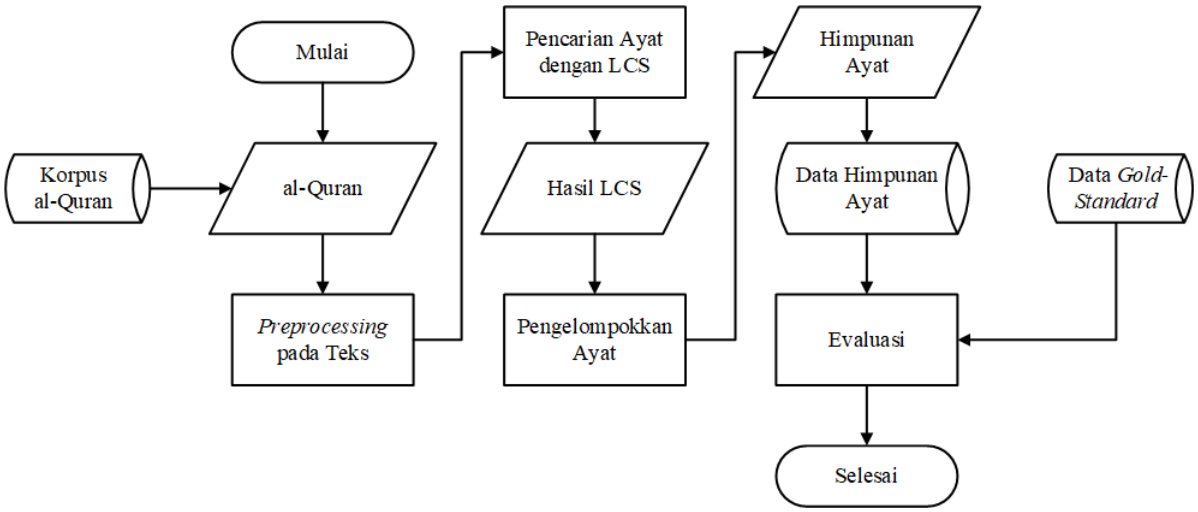
Penelitian selanjutnya yang ditemukan adalah pada tahun 2017 oleh Chaerul Hadi, et al. tentang pencarian ayat Alquran berdasarkan tema tertentu dengan menggunakan perangkat bergerak (*mobile application*) yang dibuatkan menjadi suatu aplikasi berbasis Android [9]. Pencarian ayat yang dilakukan pada penelitian ini adalah menggunakan metode *Cosine Similarity* karena metode ini dinilai memiliki nilai akurasi kemiripan yang lebih tinggi. Pada *paper* ini dijelaskan bahwa Nurdiana, et al. [16] telah melakukan penelitian berupa pencocokkan kata terhadap terjemahan Alquran dengan menggunakan metode *Cosine Similarity*, metode *Jaccard Similarity*, dan metode *k-Nearest Neighbor* (k-NN) sekaligus membandingkan hasil dari ketiga metode tersebut. Dari hasil percobaan yang dilakukan sebanyak 33 kali dengan *key* yang berbeda terhadap 6326 dokumen, metode *Cosine Similarity* mendapatkan nilai kemiripan yang paling tinggi dengan nilai 41% dibandingkan metode *Jaccard Similarity* dengan nilai 19% dan metode k-NN dengan nilai 40%.

Dari beberapa penelitian yang telah ditemukan, terdapat perkembangan dalam hal cara untuk mendapatkan ayat Alquran yang diinginkan oleh pengguna. Seperti dengan menggunakan kata atau frase, pelafalan kata, hingga dibuatkan suatu aplikasi berbasis Android dengan tujuan yang sama, yaitu untuk mempermudah orang-orang dalam mencari ayat yang diinginkan di Alquran. Kekurangan dari penelitian tersebut adalah sistem yang dibangun hanya menyajikan hasil pencarian ayat yang diinginkan oleh pengguna saja tanpa disimpan untuk digunakan di lain waktu. Sehingga setiap kali pengguna ingin mencari ayat yang diinginkan, sistem akan kembali melakukan proses pencarian ayat dari dokumen yang digunakan, sedangkan dokumen yang digunakan tersebut tidaklah sedikit. Dengan kata lain, kinerja sistem dalam melakukan pencarian ayat tersebut akan terasa berat dan membutuhkan waktu yang relatif lama. Oleh karena itu, hasil pencarian ayat yang telah dilakukan tersebut lebih baik disimpan dan dikumpulkan menjadi korpus agar pada pencarian ayat yang dilakukan selanjutnya oleh sistem dapat terasa lebih ringan dan membutuhkan waktu yang relatif lebih cepat.

III. PERANCANGAN SISTEM

A. *Desain Sistem*

Desain alur dalam membangun korpus dari rangkaian kata yang berulang pada Alquran dapat digambarkan secara umum seperti pada Gambar 1.



Gambar 1. Desain Alur Sistem

Desain tersebut dapat terbagi menjadi 4 tahapan. Tahapan yang pertama adalah melakukan *Text Preprocessing*, tahapan yang kedua adalah melakukan pencarian ayat yang memiliki struktur kata yang sama, tahapan yang ketiga adalah melakukan pengelompokkan ayat berdasarkan rangkaian kata terpanjang bersama pada ayat tersebut, dan tahapan yang terakhir adalah melakukan evaluasi antara data himpunan ayat dengan data *gold-standard* dengan menggunakan perhitungan *precision*, *recall*, dan *f1-score*.

B. Data yang Digunakan

1) Dataset: Dataset yang digunakan adalah seluruh ayat pada Alquran yang diambil dari *website* The Quranic Arabic Corpus[1] yang selanjutnya akan disebut sebagai **Data Arab** dan seluruh ayat terjemahan Alquran berbahasa Indonesia yang diambil dari *website* Tanzil-Quran Navigator[2] yang selanjutnya akan disebut sebagai **Data Indonesia**. Contoh Data Arab adalah seperti pada Tabel I dan contoh Data Indonesia adalah seperti pada Tabel II.

Tabel I
CONTOH DATA ARAB

Lokasi	Bentuk	Tag	Fitur
(1:1:1:1)	*bi*	P	PREFIX | *bi+*
(1:1:1:2)	*somi*	N	STEM | POS:N | LEM:*{som* | ROOT:*smw* | M | GEN
(1:1:2:1)	*{llāhi*	PN	STEM | POS:PN | LEM:*{llāh* | ROOT:*Alh* | GEN
(1:1:3:1)	*{l*	DET	PREFIX | *Al+*
(1:1:3:2)	*rāHoma'ni*	ADJ	STEM | POS:ADJ | LEM:*rāHoma'n* | ROOT:*rHm* | MS | GEN
(1:1:4:1)	*{l*	DET	PREFIX | *Al+*
(1:1:4:2)	*rāHiymi*	ADJ	STEM | POS:ADJ | LEM:*rāHiym* | ROOT:*rHm* | MS | GEN
...	...	...	...

Tabel II
CONTOH DATA INDONESIA

Surah	Ayat	Isi Ayat
1	1	*Dengan menyebut nama Allah Yang Maha Pemurah lagi Maha Penyayang.*
1	2	*Segala puji bagi Allah, Tuhan semesta alam.*
1	3	*Maha Pemurah lagi Maha Penyayang.*
1	4	*Yang menguasai di Hari Pembalasan.*
1	5	*Hanya Engkaulah yang kami sembah, dan hanya kepada Engkaulah kami meminta pertolongan.*
1	6	*Tunjukillah kami jalan yang lurus.*
1	1	*(yaitu) Jalan orang-orang yang telah Engkau beri nikmat kepada mereka; bukan (jalan) mereka yang dimurkai dan bukan (pula jalan) mereka yang sesat.*
...	...	...

Tabel I merupakan contoh Data Arab yang terdiri dari empat kolom, yaitu kolom Lokasi, Bentuk, Tag, dan Fitur. Kolom Lokasi berisikan nomor surah, nomor ayat, nomor kata, dan nomor posisi dari kata tersebut. Kolom Bentuk berisikan bentuk dari data tersebut. Kolom Tag berisikan kelas kata (*Part of Speech*) dari data tersebut, seperti N (*Noun*), V (*Verb*), ADJ (*Adjective*), ADV (*Adverb*), PN (*Pronoun*), CONJ (*Conjunction*), P (*Preposition*), dan INT (*Interjection*) [17]. Dan kolom Fitur berisikan keterangan-keterangan dari data tersebut. Contohnya, data pertama pada Tabel I berdasarkan kolom Lokasi, data ini berada di surah pertama, ayat pertama, kata pertama, dan berada di posisi pertama dari kata tersebut. Berdasarkan kolom Bentuk, data ini berbentuk *bi*. Berdasarkan kolom Tag, data ini memiliki kelas kata P, yaitu kata depan atau preposisi (*Preposition*). Dan berdasarkan kolom Fitur, terdapat keterangan PREFIX yang artinya data ini merupakan bentuk prefiks dan keterangan *bi+* yang artinya data ini menyambung dengan data berikutnya. Contoh selanjutnya, data kedua pada Tabel I berdasarkan kolom Lokasi, data ini berada di surah pertama, ayat pertama, kata pertama, dan berada di posisi kedua dari kata tersebut.

[1]Kais Dukes, "The Quranic Arabic Corpus", http://corpus.quran.com/ (diakses pada 25 April 2019, pukul 10.24)
[2]Hamid Zarrabi-Zadeh, "Tanzil-Quran Navigator", http://tanzil.net/ (diakses pada 25 April 2019, pukul 11.05)

Berdasarkan kolom bentuk, data ini berbentuk *somi*. Berdasarkan kolom Tag, data ini memiliki kelas kata N, yaitu kata benda (*Noun*). Dan berdasarkan kolom Fitur, terdapat keterangan STEM yang artinya data ini adalah hasil *stemming*, keterangan POS:N yang artinya data ini memiliki kelas kata yaitu kata benda (*Noun*), keterangan LEM:*{som* yang artinya data ini memiliki bentuk *Lemma* yaitu *{som*, keterangan ROOT:*smw* yang artinya data ini memiliki bentuk dasar yaitu *smw*, keterangan M atau *Masculine* yang artinya data ini merupakan sebuah kata yang mengacu kepada kata-kata maskulin atau bersifat kelelakian berdasarkan tata bahasa arab [3], dan keterangan GEN atau *Genitive* yang artinya data ini menunjukkan jenis kata kepunyaan atau kepemilikan [13]. Sedangkan Tabel II merupakan contoh Data Indonesia yang terdiri dari tiga kolom, yaitu kolom Surah, Ayat, dan Isi Ayat. Kolom Surah berisikan nomor surah, kolom Ayat berisikan nomor ayat, dan kolom Isi Ayat berisikan isi dari ayat tersebut. Contohnya, data pertama pada Tabel II berdasarkan kolom Surah, data ini berada di surah pertama. Berdasarkan kolom Ayat, data ini berada di ayat pertama. Dan berdasarkan kolom Isi Ayat, data ini berisi yaitu *Dengan menyebut nama Allah Yang Maha Pemurah lagi Maha Penyayang*.

Khusus untuk Data Arab, data yang diperoleh dari *website* The Quranic Arabic Corpus[3] adalah berupa Aksara Romawi atau Aksara Latin. Data dengan bentuk inilah yang akan digunakan dalam komputasi pada sistem, karena aksara ini merupakan aksara yang paling banyak digunakan di dunia [1]. Sedangkan data dengan Aksara Arab pada pembahasan selanjutnya merupakan hasil *encode* penulis secara manual untuk mempermudah pembaca dalam memahami penelitian ini.

2) Data Gold-Standard: Data *gold-standard* yang digunakan adalah data yang dibuat oleh penulis dengan jumlah 789 ayat atau data Alquran dari awal surah pertama yaitu Surah *al-Fatihah* hingga akhir surah kelima yaitu Surah *al-Maidah* dengan ketentuan bahwa ayat yang diprediksi sama dengan ayat lainnya adalah ayat yang memiliki rangkaian kata terpanjang dengan jumlah rangkaian kata lebih dari atau sama dengan selisih antara jumlah kata pada ayat pembanding dan jumlah rangkaian kata terpanjang tersebut, seperti pada Rumus 1. Hal tersebut dikarenakan penulis berasumsi bahwa jumlah rangkaian kata terpanjang yang baik adalah yang jumlah katanya sama dengan atau lebih dominan dibandingkan sisa kata dari teks tersebut. Jumlah data yang didapat pada saat membuat data *gold-standard* adalah seperti pada Tabel III.

$$lcs >= (n - lcs) \tag{1}$$

Dengan:
lcs: jumlah rangkaian kata terpanjang
n: jumlah keseluruhan kata pada ayat pembanding

Tabel III
JUMLAH DATA GOLD-STANDARD

Jenis Data	Pasangan Ayat	Himpunan Ayat
Data Arab	1180	613
Data Indonesia	1529	537

Pada Tabel III terdapat perbedaan jumlah pasangan ayat dan jumlah himpunan ayat antara Data Arab dan Data Indonesia, sedangkan jumlah ayat yang digunakan untuk kedua data tersebut sama yaitu sebanyak 789 ayat. Hal ini dikarenakan, *Text Preprocessing* yang digunakan untuk membuat data *gold-standard* antara Data Arab dan Data Indonesia pun berbeda. *Text Preprocessing* yang digunakan untuk membuat data *gold-standard* dari Data Arab dan Data Indonesia sama dengan yang digunakan di dalam sistem pada penelitian ini dan akan dibahas di subbab berikutnya.

Adapun contoh data *gold-standard* pada Data Arab yang didapat seperti pada Tabel IV dan pada Data Indonesia yang didapat seperti pada Tabel V.

C. Text Preprocessing

Text Preprocessing merupakan proses merubah data yang digunakan sebagai objek penggalian menjadi data yang terstruktur [11]. Pada Data Arab, data yang berbentuk prefiks dan sufiks akan dihilangkan dan

[3]Kais Dukes, "The Quranic Arabic Corpus", http://corpus.quran.com/ (diakses pada 25 April 2019, pukul 10.24)

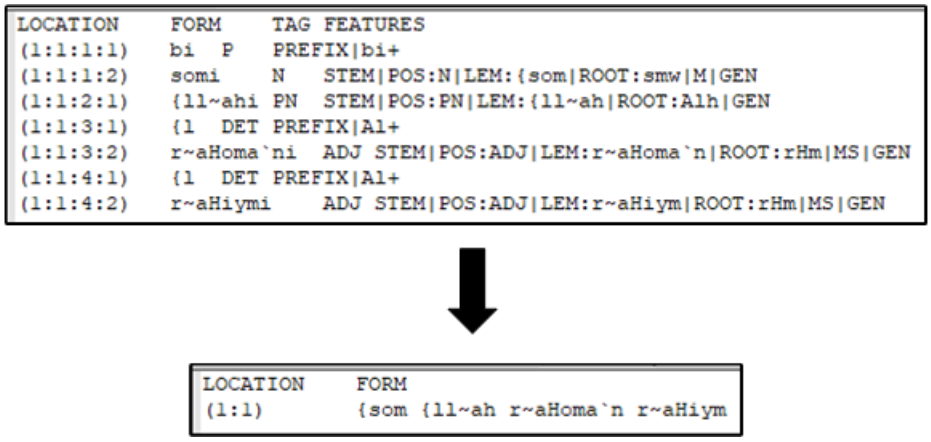
Tabel IV
CONTOH GOLD-STANDARD DATA ARAB

Rangkaian Kata Terpanjang	Surah dan Ayat	Jumlah Ayat	Jumlah Kata
امَنَ مَا أَنزَلَ عَلَى مَا أَنزَلَ مِن قَبلَ أَخَر	(2:4), (4:162)	2	9
ذَالِكَ بَيَّن اللهُ كُم آيَة لَعَلَّ	(2:242), (3:103), (5:89)	3	6
أَيُّهَا الَّذِي امَنَ مِن اللهُ إِن كَانَ	(2:172), (4:59), (5:106), (5:57)	4	7
جَنَّة جَرَي مِن تَهتَ نَهَر خَالِد فِي	(3:136), (3:198), (4:57), (4:122), (5:85)	5	7
قَاتَلَ فِي سَبِيل اللهُ اللهُ	(2:190), (2:244), (3:13), (3:146), (4:74), (4:84)	6	5
...	...	...	...

Tabel V
CONTOH GOLD-STANDARD DATA INDONESIA

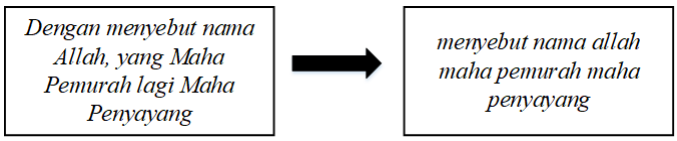
Rangkaian Kata Terpanjang	Surah dan Ayat	Jumlah Ayat	Jumlah Kata
allah bumi allah maha pengampun maha penyayang	(3:129), (4:100)	2	7
hai bani israil ingatlah nikmat anugerahkan kepadamu	(2:40), (2:47), (2:122)	3	7
allah sesungguhnya allah maha maha penyayang	(2:37), (2:143), (4:106), (5:98)	4	6
sesungguhnya allah maha mengetahui maha bijaksana	(4:11), (4:24), (4:104), (4:111), (4:170)	5	6
allah sesungguhnya allah maha pengampun maha penyayang	(2:173), (2:199), (4:106), (5:3), (5:39), (5:98)	6	7
...	...	...	...

data sisanya akan diambil bentuk *Lemma* dan digabungkan berdasarkan surah dan ayatnya, seperti pada Gambar 2. Bagi data yang tidak memiliki bentuk *Lemma*, akan diambil bentuk data yang terdapat di kolom Bentuk pada Tabel I. Sedangkan pada Data Indonesia, dilakukan pengubahan bentuk penulisan menjadi bentuk *lowercase*, penghapusan semua karakter yang tidak dibutuhkan, dan penghapusan *stopword* yaitu kata-kata yang dianggap tidak memiliki pengaruh pada suatu teks [14], seperti pada Gambar 3.



Gambar 2. Contoh Hasil *Text Preprocessing* Data Arab

Kemudian menentukan dan merubah format penulisan data dengan menggabungkan nomor surah dan



Gambar 3. Contoh Hasil *Text Preprocessing* Data Indonesia

nomor ayat agar Data Arab dan Data Indonesia memiliki format yang sama, yaitu terdiri dari kolom Surah dan Ayat dan kolom Isi Ayat, sehingga dapat memudahkan untuk diproses di tahapan selanjutnya. *Text Preprocessing* ini pun digunakan dalam membangun data *gold-standard* untuk Data Arab dan Data Indonesia. Contohnya seperti penulisan isi dari Surah *al-Fatihah* pada Tabel VI dan Tabel VII.

Tabel VI
FORMAT PENULISAN DATA ARAB

Surah dan Ayat	Isi Ayat
(1:1)	بِ اسم اللّٰه رَحمٰن رَحِيم
(1:2)	حَمَد اللّٰه رَبّ عٰلَمِين
(1:3)	رَحمٰن رَحِيم
(1:4)	مٰلِك يَوم دِين
(1:5)	إِيَّا عَبَدَ إِيَّا استَعِين
(1:6)	هُدَى صِّرَاط مُستَقِيم
(1:7)	صِرَاط الَّذِينَ أَنعَم عَلَى غَير مَغضُوب عَلَى لَا ضَالِّين
...	...

Tabel VII
FORMAT PENULISAN DATA INDONESIA

Surah dan Ayat	Isi Ayat
(1:1)	*menyebut nama allah maha pemurah maha penyayang*
(1:2)	*puji allah tuhan semesta alam*
(1:3)	*maha pemurah maha penyayang*
(1:4)	*menguasai hari pembalasan*
(1:5)	*engkaulah sembah engkaulah meminta pertolongan*
(1:6)	*tunjukillah jalan lurus*
(1:7)	*jalan orang orang engkau beri nikmat dimurkai jalan sesat*
...	...

D. Pencarian Ayat yang Sama

Tahapan ini adalah mencari ayat yang sama dengan menggunakan pendekatan *Longest Common Subsequence* (LCS). Metode LCS ini merupakan metode pencarian suatu urutan atau rangkaian (*sequence*) terpanjang bersama dari suatu set urutan [7]. Selain mirip dengan rangkaian *substring*, metode ini pun dapat mendeteksi rangkaian yang pada pertengahan rangkaian tersebut terdapat beberapa data yang terselip diantaranya. Contohnya seperti data 1 yaitu "ABCDEFG" dan data 2 yaitu "ABDECFG" yang memiliki rangkaian terpanjang bersama yaitu "ABDEFG" [6]. Tahapan ini dimaksudkan untuk menemukan ayat yang sama berdasarkan rangkaian kata terpanjang bersama yang digunakan dalam menyusun ayat tersebut dengan cara memotong rangkaian tersebut menjadi beberapa kata berdasarkan karakter spasi ( ).

Contohnya seperti pada pasangan ayat berikut:

- Contoh pada Data Arab:
  - Ayat 1 pada (QS:2:48):

    اِتَّقُ يَوم لَا جَزَى نَفس عَن نَفس شَيءَ لَا يَقَبَّل مِن شَفَاع لَا أَخَذَ مِن عَدل لَا هُم نَصَرَ
  - Ayat 2 pada (QS:2:123):

    اِتَّقُ يَوم لَا جَزِي نَفس عَن نَفس شَيءَ لَا يَقَبَّل مِن هَا عَدل لَا نَفَعَ هَا شَفَاع لَا هُم نَصَرَ
  - Hasil rangkaian terpanjang:

    اِتَّقُ يَوم لَا جَزَى نَفس عَن نَفس شَيءَ لَا يَقَبَّل مِن عَدل لَا هُم نَصَرَ
- Contoh pada Data Indonesia:
  - Ayat 1 pada (QS:2:136):
    hai mukmin ***beriman allah diturunkan diturunkan ibrahim ismail ishaq yaqub cucunya musa isa nabi*** *nabi tuhannya* ***membeda bedakan*** *tunduk patuh*
  - Ayat 2 pada (QS:3:84):
    beriman allah diturunkan diturunkan ibrahim ismail ishaq yaqub *anaknya* ***musa isa nabi*** *tuhan* ***membeda bedakan*** *menyerahkan*
  - Hasil rangkaian terpanjang:
    beriman allah diturunkan diturunkan ibrahim ismail ishaq yaqub musa isa nabi membeda bedakan

Dari tahapan ini, dapat diperoleh pasangan-pasangan ayat yang sama berdasarkan rangkaian kata terpanjang bersama pada Alquran, seperti pada Tabel VIII untuk Data Arab dan Tabel IX untuk Data Indonesia.

Tabel VIII
CONTOH HASIL LCS PADA DATA ARAB

Hasil LCS	QS-1	QS-2	Jumlah Kata
ذٰلِكَ بَيَان آيَة	(2:118)	(2:242)	3
اَلَّذِي آمَنَ اقَامَ صَلَاة	(2:3)	(2:777)	4
اَللّٰه اتَّبَعَ مِلَّة إِبرَاهِيم حَنِيف	(3:95)	(4:125)	5
امَنَ مَا أُنزَلَ مَا أُنزَلَ مِن	(2:4)	(3:84)	6
اَلَّذِي بَخِلَ مَا آتَى اللّٰه مِن فَضَل	(3:180)	(4:37)	7
اَلَّذِي لَا اعتَدَى إِنَّ اللّٰه احبَب مُعتَدِين	(2:190)	(5:87)	8
اَللّٰه مَا غَفَرَ مَن شَآءَ عِذَّبَ مَن شَآءَ اللّٰه	(5:18)	(3:129)	9
اَللّٰه مَا فِي سَمَاء مَا فِي أَرَض اللّٰه كُلُّ شَيءَ	(2:284)	(4:126)	10
إِذ أَخَذَ مِيثَق رَفَعَ فَوقَ طُور أَخَذَ مَا آتَى قُوَّة مَا	(2:93)	(2:63)	11
اَلَّذِي اللّٰه هُم أَجر عِندَ رَبّ لَا خَوف عَلَى لَا هُم يَحزُن	(2:62)	(2:262)	12
...	...	...	...

Hasil dari pencarian ayat yang sama dengan menggunakan sistem yang dibangun pada penelitian ini terdapat hasil yang berbeda antara Data Arab dan Data Indonesia. Contohnya pada Tabel VIII, QS (2:118) dengan QS (2:242) pada Data Arab diprediksi sama oleh sistem tetapi belum tentu dapat ditemukan pada

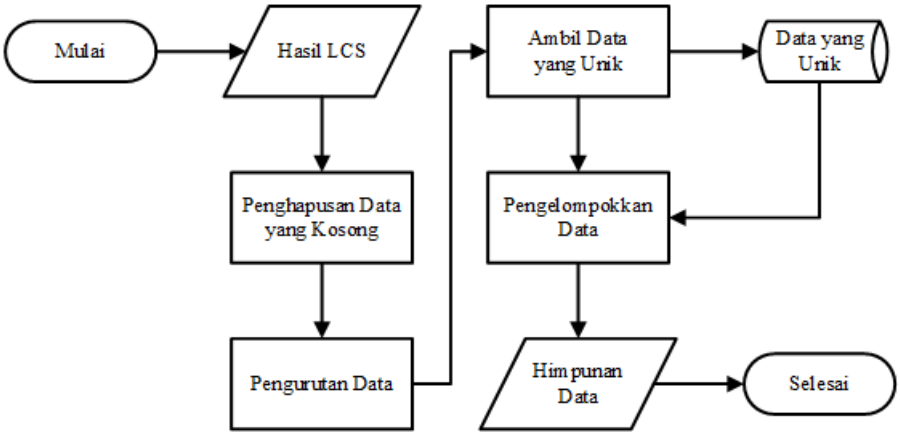
Tabel IX
CONTOH HASIL LCS PADA DATA INDONESIA

Hasil LCS	QS-1	QS-2	Jumlah Kata
allah dikembalikan urusan	(2:210)	(3:109)	3
ahli kitab allah mengetahui	(2:102)	(3:70)	4
kepunyaan allah langit bumi allah	(3:180)	(3:189)	5
sesungguhnya allah maha pengampun maha penyayang	(2:173)	(2:192)	6
kecuali sesungguhnya allah maha pengampun maha penyayang	(4:23)	(3:89)	7
allah menentukan rahmat kenabian allah mempunyai karunia besar	(2:105)	(3:74)	8
mohonkanlan tuhanmu untuk menerangkan sapi betina sesungguhnya sapi sapi	(2:68)	(2:70)	9
sesungguhnya allah kerajaan langit bumi dikehendaki allah maha kuasa atas	(5:17)	(5:40)	10
sesungguhnya penciptaan langit bumi silih bergantinya malam siang terdapat tanda tanda	(2:164)	(3:190)	11
hai ahli kitab sesungguhnya datang rasul menjelaskan kepadamu sesungguhnya datang kepadamu allah	(5:19)	(5:15)	12
...	...	...	...

hasil LCS pada Data Indonesia. Begitu pun sebaliknya, pada Tabel IX, QS (2:210) dengan QS (3:109) pada Data Indonesia diprediksi sama oleh sistem tetapi belum tentu dapat ditemukan pada hasil LCS Data Arab. Seperti yang telah dijelaskan pada subbab sebelumnya, bahwa *Text Preprocessing* yang digunakan pada Data Arab berbeda dengan *Text Preprocessing* yang digunakan pada Data Indonesia sehingga data yang diperoleh pun berbeda.

E. Pengelompokkan Ayat

Tahapan ini adalah mengelompokkan ayat-ayat yang diprediksi sama berdasarkan hasil LCS yang diperoleh dari sistem. Adapun alur proses dari pengelompokkan ayat dimulai dari penghapusan data yang kosong yaitu data yang tidak memiliki rangkaian kata bersama. Kemudian data sisanya akan diurutkan berdasarkan hasil LCS dan diambil hasil LCS yang unik untuk dijadikan sebagai rujukan dari proses pengelompokkan data. Alur dari proses ini dapat dilihat pada Gambar 4.



Gambar 4. Alur Proses Pengelompokkan Ayat

Tahapan ini akan menghasilkan suatu himpunan data dari ayat-ayat yang memiliki rangkaian kata terpanjang bersama yang terhimpun berdasarkan hasil LCS, seperti pada Tabel X dan Tabel XI.

Sama seperti pada pembahasan sebelumnya bahwa hasil yang didapat dari Data Arab akan berbeda dengan hasil yang didapat dari Data Indonesia. Hal tersebut dikarenakan penggunaan *Text Preprocessing* yang berbeda antara Data Arab dan Data Indonesia.

Tabel X
CONTOH HASIL PENGELOMPOKKAN DATA ARAB

Rangkaian Kata Terpanjang	Surah dan Ayat	Jumlah Ayat	Jumlah Kata
اَللّٰه عٰلَمِين	(1:2), (2:251), (3:108), (3:33), (3:42), (3:97), (5:20), (5:115)	8	2
اَللّٰه غَفُور رَحِيم	(2:192), (3:129), (4:96), (4:100), (4:110), (4:152), (5:34), (5:74), (5:98)	9	3
اَلَّذِي كَفَرَ أُلَٰئِكَ أَصحَاب	(2:257), (3:116), (5:10), (5:86)	4	4
ذٰلِك بَيَّان اللّٰه آيَة لَعَلَّى	(2:187), (2:219), (2:242), (2:266)	4	5
اَلَّذِي إِذَا أَصحَاب مُصِيبَة اللّٰه إِنَّ	(2:156), (5:106)	2	6
إِنَّ الَّذِي اشتَرَ اللّٰه هُم عَذَاب أَلِيم	(2:174), (3:177), (3:77)	3	7
اَلَّذِي كَفَرَ أُلَٰئِكَ أَصحَاب نَار هُم فِي خَالِد	(2:39), (2:257), (3:116)	9	8
اَللّٰه مُلك سَمَاء أَرَض اللّٰه عَلَى كُلُّ شَيءَ قَدِير	(3:189), (5:17), (5:40)	3	9
يَوم ثُمَّ وَفّىء كُلُّ نَفس مَا كَسَبَ هُم لَا ظَلَمَ	(2:281), (3:161)	2	10
بُنَي إِسرَائِيلَ اذكَرَ نِعمَة الَّذِي أَنعَمَ عَلَى أَنّ فَضَّلَ عَلَى عٰلَمِين	(2:47), (2:122)	2	11
...	...	...	...

F. Evaluasi

Evaluasi merupakan tahapan akhir yang dilakukan untuk mengetahui performansi sistem yang digunakan. Umumnya, menggunakan perhitungan *precision*, *recall*, dan *f1-score* [18]. Tabel *Confusion Matrix* dapat membantu dalam memperoleh nilai dari perhitungan tersebut karena tabel ini berisi data prediksi yang positif dan negatif yang dihasilkan oleh sistem, dan data aktual yang positif dan negatif di dunia nyata [8]. Untuk mendapatkan hasil performansi pada penelitian ini, dilakukan perbandingan antara data yang dihasilkan oleh sistem dengan data *gold-standard* yang telah dibuat oleh penulis dengan menggunakan Rumus 1.

Pada Tabel XII terdapat perhitungan 4 kategori dengan menggunakan ketentuan kelas positif dan negatif pada sistem dan aktual yang positif dan negatif. Kelas positif pada sistem menandakan bahwa ayat tersebut diprediksi sama dengan ayat lainnya oleh sistem. Sebaliknya, kelas negatif pada sistem menandakan bahwa ayat tersebut tidak diprediksi sama dengan ayat lainnya oleh sistem. Sedangkan aktual yang positif menandakan bahwa data tersebut terdapat pada data *gold-standard*. Sebaliknya, aktual yang negatif menandakan bahwa data tersebut tidak terdapat pada data *gold-standard*. Perhitungan 4 kategori tersebut terdiri dari:

1) TP (*True Positive*), yaitu perhitungan dari kelas positif sistem dan aktual yang positif.
2) TN (*True Negative*), yaitu perhitungan dari kelas negatif sistem dan aktual yang negatif.
3) FP (*False Positive*), yaitu perhitungan dari kelas positif sistem dan aktual yang negatif.
4) FN (*False Negative*), yaitu perhitungan dari kelas negatif sistem dan aktual yang positif.

Perhitungan yang umum digunakan untuk memperoleh performansi adalah sebagai berikut:

Tabel XI
CONTOH HASIL PENGELOMPOKKAN DATA INDONESIA

Rangkaian Kata Terpanjang	Surah dan Ayat	Jumlah Ayat	Jumlah Kata
hari pembalasan	(1:4), (3:161)	2	2
sesungguhnya allah bumi	(3:5), (3:137), (4:97), (4:170), (4:171), (5:33), (5:40), (5:97)	8	3
kepunyaan allah langit bumi	(2:255), (2:284), (3:109), (3:129), (3:180), (3:189), (4:126), (4:131), (4:132), (5:120)	10	4
allah sesungguhnya allah maha mengetahui	(2:244), (2:282), (2:283), (3:154), (4:32), (4:135), (5:8), (5:97)	8	5
allah menghendaki niscaya umat allah berbuat	(4:133), (5:48)	2	6
allah sesungguhnya allah maha pengampun maha penyayang	(2:173), (2:199), (4:106), (5:3), (5:39), (5:98)	6	7
katakanlah allah langit bumi allah maha kuasa atas	(3:29), (5:17)	2	8
orang berbuat dosa sesungguhnya allah maha pengampun maha penyayang	(2:182), (5:3)	2	9
sesungguhnya allah kerajaan langit bumi dikehendaki allah maha kuasa atas	(5:17), (5:40)	2	10
sesungguhnya allah mengampuni dosa mengampuni dosa syirik dikehendaki barangsiapa mempersekutukan allah	(4:116), (4:48)	2	11
...	...	...	...

Tabel XII
TABEL CONFUSION MATRIX [8]

	Positif (Aktual)	Negatif (Aktual)
Positif (Sistem)	TP	FP
Negatif (Sistem)	FN	TN

1) *Precision*
Precision adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem [19].

$$Precision = \frac{TP}{TP + FP} \times 100\% \tag{2}$$

2) *Recall*
Recall merupakan tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi [19].

$$Recall = \frac{TP}{TP + FN} \times 100\% \tag{3}$$

3) *F1-Score*
F1-Score merupakan pengukuran performansi yang menggabungkan perhitungan *precision* dan *recall* [19].

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \times 100\% \tag{4}$$

IV. HASIL DAN PEMBAHASAN

Pada bagian ini akan dilakukan pengujian terhadap sistem yang telah dibuat dengan menggunakan jumlah Data Arab dan Data Indonesia yang sama. Kemudian hasil tersebut dibandingkan dengan data *gold-standard* untuk mendapatkan hasil performansi sistem yang optimal.

A. Skenario Pengujian

Pengujian dilakukan dengan menggunakan Data Arab dan Data Indonesia yang dimasukkan ke dalam sistem adalah sebanyak 789 ayat, lebih tepatnya adalah data dari awal surah pertama yaitu Surah *al-Fatihah* sampai akhir surah kelima yaitu Surah *al-Maidah*. Kemudian dilakukan pencarian ayat yang sama untuk mendapatkan rangkaian kata terpanjang bersama dengan menggunakan pendekatan *Longest Common Subsequence* (LCS) dan menggunakan ketentuan seperti pada Rumus 5 dengan nilai i adalah 1, 2, dan 3 untuk Data Arab dan Data Indonesia. Ketentuan ini digunakan untuk menyaring data agar data yang diambil oleh sistem adalah data dengan jumlah kata yang dominan dari pada kata sisanya pada ayat tersebut. Tabel VIII dan Tabel IX adalah contoh hasil pencarian ayat yang sama yang didapat dari sistem.

$$(lcs >= i) \wedge (a <= 1) \tag{5}$$

Dengan:
lcs: jumlah rangkaian kata terpanjang
i: jumlah kata
a: hasil dari $n - lcs$
n: jumlah keseluruhan kata pada ayat pembanding

Hasil dari pencarian ayat yang sama tersebut kemudian dikumpulkan berdasarkan rangkaian kata terpanjang bersama yang diperoleh. Tabel X dan Tabel XI adalah contoh hasil pengelompokkan data yang didapat dari sistem. Data akhir tersebut dibandingkan dengan data *gold-standard* yang dibuat oleh penulis menggunakan paremeter seperti pada Rumus 1. Pengujian dilanjutkan untuk mendapatkan nilai *precision*, *recall*, dan *f1-score* dari sistem yang sudah dibangun.

B. Analisis Hasil Pengujian

Adapun hasil pencarian ayat yang sama yang didapat dari sistem kembali ditemukan perbedaan, yaitu ayat yang diprediksi sama pada Data Arab belum tentu sama dengan ayat yang diprediksi sama pada Data Indonesia. Sehingga jumlah pasangan ayat dan himpunan ayat yang diperoleh ketika melakukan pengelompokkan data berdasarkan rangkaian kata terpanjang bersama, akan berbeda pula. Hal tersebut telah dijelaskan dipembahasan sebelumnya, yaitu dikarenakan kedua data ini menggunakan *Text Preprocessing* yang berbeda antara Data Arab dan Data Indonesia.

Adapun hasil dari pengujian yang telah dilakukan adalah beragam karena bergantung dari nilai i yang digunakan, seperti pada Tabel XIII. Setelah itu, hasil dari sistem dibandingkan dengan data *gold-standard* untuk mendapatkan nilai evaluasi dengan menggunakan perhitungan *precision*, *recall*, dan *f1-score*. Hasil perhitungan yang telah dilakukan terhadap Data Arab seperti pada Tabel XIV dan untuk Data Indonesia seperti pada Tabel XV.

Tabel XIII
HASIL PENGUJIAN

Jenis Data	**Nilai i**	**Pasangan Ayat**	**Himpunan Ayat**
Data Arab	1	2562	1091
	2	2426	1084
	3	2280	1065
Data Indonesia	1	4227	902
	2	3984	893
	3	3004	856

Tabel XIV
HASIL EVALUASI DATA ARAB

Nilai i	***True Positive* (TP)**	***Precision* (%)**	***Recall* (%)**	***F1-Score* (%)**
1	1180	46.06	100	63.07
2	1148	47.32	97.29	63.67
3	1075	47.15	91.10	62.14
Rata-Rata		**46.84**	**96.13**	**62.96**

Tabel XV
HASIL EVALUASI DATA INDONESIA

Nilai i	***True Positive* (TP)**	***Precision* (%)**	***Recall* (%)**	***F1-Score* (%)**
1	1529	36.17	100	53.13
2	1529	38.38	100	55.47
3	1417	47.17	92.67	62.52
Rata-Rata		**40.57**	**97.56**	**57.04**

Dari hasil perhitungan di atas, didapat hasil yang beragam karena nilai i dari Rumus 5 pun beragam. Untuk Data Arab, sesuai dengan Tabel XIV, hasil terbaik berdasarkan nilai *f1-score* adalah ketika nilai i sama dengan 2, yaitu nilai *precision* sebesar 47.32%, nilai *recall* sebesar 97.29%, dan nilai *f1-score* sebesar 63.67%. Sedangkan untuk Data Indonesia, sesuai dengan Tabel XV, hasil terbaik berdasarkan nilai *f1-score* adalah ketika nilai i sama dengan 3, yaitu nilai *precision* sebesar 47.17%, nilai *recall* sebesar 92.67%, dan nilai *f1-score* sebesar 62.52%.

V. KESIMPULAN

Penelitian ini dilakukan dengan tujuan untuk membangun korpus dari rangkaian ayat yang berulang pada Alquran diawali dengan melakukan pencarian ayat yang sama berdasarkan rangkaian kata terpanjang bersama yang digunakan dalam menyusun ayat tersebut. Pencarian ayat yang sama ini dilakukan dengan menggunakan pendekatan *Longest Common Subsequence* (LCS) untuk mendapatkan rangkaian terpanjang bersama dengan ayat lainnya dan ditambah dengan ketentuan sesuai dengan Rumus 5 dengan nilai i adalah 1, 2, dan 3. Ketentuan ini digunakan untuk menyaring data agar data yang diambil adalah data yang memiliki jumlah rangkaian kata terpanjang yang dominan pada ayat pembanding. Kemudian data tersebut dikumpulkan menjadi korpus. Data pada korpus tersebut kemudian dibandingkan dengan data *gold-standard* dan dievaluasi dengan menggunakan perhitungan *precision*, *recall*, dan *f1-score*.

Dari hasil evaluasi, didapat nilai terbaik berdasarkan nilai *f1-score* untuk Data Arab dan Data Indonesia. Hasil terbaik pada Data Arab adalah nilai *precision* sebesar 47.32%, nilai *recall* sebesar 97.29%, dan nilai *f1-score* sebesar 63.67%. Sedangkan untuk Data Indonesia adalah nilai *precision* sebesar 47.17%, nilai *recall* sebesar 92.67%, dan nilai *f1-score* sebesar 62.52%. Jadi dapat disimpulkan bahwa sistem yang dibangun pada penelitian ini menghasilkan data yang masih layak untuk dijadikan korpus dari rangkaian kata yang berulang pada ayat Alquran. Nilai *recall* yang didapat lebih dari 90% adalah bukti bahwa sistem ini dapat memperoleh hampir semua data yang terdapat pada data *gold-standard*, walaupun sistem ini masih dapat memprediksi banyak ayat yang sama selain yang terdapat pada data *gold-standard*. Hal ini dibuktikan oleh nilai *precision* yang kurang dari 50%.

Saran untuk pengembangan selanjutnya adalah penggunaan rumus yang lebih sesuai dalam memprediksi ayat yang sama dengan ayat lainnya agar akurasi dari sistem yang dibangun dapat lebih optimal dengan bukti nilai *precision* yang mendekati 100%.

Pustaka

[1] W Sidney Allen and William Sidney Allen. *Vox Latina - The Pronunciation of Classical Latin*. Cambridge University Press, 1989.
[2] Agus Sofiyan Anwar, Zainal Abidin, and Ririen Kusumawati. Mesin Pencari Ayat Al Quran Menggunakan Inexact String Matching. *MATICS*, 2012.
[3] Moch Anwar. Ilmu Nahwu Terjemahan Matan Al Jurumiyah dan Imrithy Berikut Penjelasannya. *Revisi Edisi*, 2, 2009.
[4] Syed Uzair Aqeel, Steve Beitzel, Eric Jensen, David Grossman, and Ophir Frieder. On The Development of Name Search Techniques for Arabic. *Journal of the American Society for Information Science and Technology*, 57(6):728–739, 2006.
[5] Nashruddin Baidan. *Metode Penafsiran Al-Quran: Kajian Kritis terhadap Ayat-Ayat yang Beredaksi Mirip*. Pustaka Pelajar, 2002.
[6] Lasse Bergroth, Harri Hakonen, and Timo Raita. A Survey of Longest Common Subsequence Algorithms. In *Proceedings Seventh International Symposium on String Processing and Information Retrieval. SPIRE 2000*, pages 39–48. IEEE, 2000.
[7] Thomas H Cormen, Charles E Leiserson, Ronald L Rivest, and Clifford Stein. *Introduction to algorithms*. MIT press, 2009.
[8] Cyril Goutte and Eric Gaussier. A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. In *European Conference on Information Retrieval*, pages 345–359. Springer, 2005.
[9] Chaerul Hadi and Muhammad Rifqi Ma'arif. Implementasi Cosine Similarity dalam Aplikasi Pencarian Ayat Al-Quran Berbasis Android. *e-Journal*, 2017.
[10] Muhammad Baqir Hakim. *Ulumul Quran*. Al-Huda, 2006.
[11] Jiawei Han, Jian Pei, and Micheline Kamber. *Data Mining: Concepts and Techniques*. Elsevier, 2011.
[12] Muhammad Abrar Istiadi. Sistem Pencarian Ayat Al-Quran Berbasis Kemiripan Fonetis, 2012.
[13] Christian Lehmann. Directions for Interlinear Morphemic Translations. *Folia linguistica*, 16(1-4):199–224, 1982.
[14] LAN Man. *A New Term Weighting Method for Text Categorization*. PhD thesis, Nasional Univeristy of Singapore, Singapore, 2007.
[15] Ahsin Sakho Muhammad. *Oase Al-Quran Penyejuk Kehidupan*. Qaf, 2017.
[16] Ogie Nurdiana, Jumadi Jumadi, and Dian Nursantika. Perbandingan Metode Cosine Similarity dengan Metode Jaccard Similarity pada Aplikasi Pencarian Terjemah Al-Quran Dalam Bahasa Indonesia. *Jurnal Online Informatika*, 1(1):59–63, 2016.
[17] Slav Petrov, Dipanjan Das, and Ryan McDonald. A Universal Part-of-Speech Tagset. *arXiv preprint arXiv:1104.2086*, 2011.
[18] Luis Torgo and Rita Ribeiro. Precision and Recall for Regression. In *International Conference on Discovery Science*, pages 332–346. Springer, 2009.
[19] Deqing Wang. *A Two-Stage Feature Selection Method for Text Categorization*. PhD thesis, Beihang University, Beijing, China, 2009.