

SISTEM Pencarian Terjemahan Ayat Al-Qur'an Menggunakan Teknik Text Mining dan Regular Expression

Muhammad Syazidan, Dadang Iskandar Mulyana, Kastum

Teknik Informatika, STIKOM Cipta Karya Informatika

Jalan Radin Inten II, Jakarta Timur, Indonesia

syazidan313@gmail.com

ABSTRAK

Meningkatnya akses terhadap teks Al-Qur'an melalui platform digital memberikan peluang untuk menghadirkan sistem pencarian ayat yang lebih cepat dan tematik. Namun demikian, sebagian besar sistem pencarian yang tersedia masih bergantung pada pencocokan literal, sehingga sering kali menghasilkan hasil yang tidak akurat atau kurang relevan karena tidak memperhatikan konteks maupun struktur kata secara utuh. Berdasarkan hal tersebut, penelitian ini bertujuan untuk merancang dan mengembangkan sistem pencarian ayat Al-Qur'an berbasis *web* yang mampu menyajikan hasil pencarian yang lebih tepat sasaran melalui pendekatan pencocokan kata kunci secara utuh. Sistem dibangun menggunakan teknik *text mining* untuk pra-pemrosesan data dan *regular expression* sebagai metode pencarian utama. Arsitektur sistem mengandalkan *framework Python Flask* di sisi *server* dan *HTML/CSS/JavaScript* untuk antarmuka pengguna. Hasil pengujian menunjukkan bahwa sistem mampu menyaring hasil pencarian dengan tingkat relevansi yang lebih tinggi dibandingkan pencarian literal, serta mendukung fitur ekspor ke dalam format *CSV* dan *PDF*. Sistem ini memberikan alternatif pencarian yang efisien, mudah diakses, dan kontekstual bagi pengguna yang ingin menelusuri ayat-ayat Al-Qur'an berdasarkan tema tertentu.

Kata kunci : Pencarian Ayat Al-Qur'an, Text Mining, Regular Expression, Whole-Word Matching, Aplikasi Web.

1. PENDAHULUAN

Di era digital saat ini, cara kita berinteraksi dengan Al-Qur'an sebagai pedoman hidup telah berubah drastis. Kitab suci yang kita junjung kini dapat diakses kapan saja melalui berbagai aplikasi dan situs *web*, tidak lagi terbatas pada *mushaf* cetak [1].

Namun, kemudahan ini membawa sebuah tantangan baru: bagaimana kita bisa menemukan ayat-ayat yang relevan dengan sebuah tema atau kata kunci secara cepat dan tepat? Bagi banyak orang, mulai dari pelajar hingga peneliti, mencari topik tertentu di tengah lautan teks Al-Qur'an menjadi tugas yang rumit jika hanya mengandalkan fitur pencarian sederhana yang seringkali kurang presisi [2].

Untuk menjembatani kesenjangan ini, teknologi modern seperti *Text Mining* hadir sebagai solusi, menawarkan cara untuk menggali informasi berharga dari data teks yang melimpah [3].

Kemampuannya sudah terbukti dalam berbagai bidang, bahkan untuk tugas rumit seperti menganalisis teks Al-Qur'an untuk deteksi hukum *tajwid*, dan perannya kini semakin diakui di tingkat global. Sebagai bagian dari *Natural Language Processing (NLP)* yang didukung oleh teknik seperti tokenisasi, pendekatan ini disempurnakan dengan *Regular Expression (Regex)* [4].

Regex berfungsi layaknya *'filter'* cerdas yang memastikan mesin pencari menemukan kata yang kita maksud secara utuh, sebuah metode yang presisinya telah teruji di berbagai aplikasi, termasuk *chatbot* [5].

Atas dasar itulah, penelitian ini berfokus pada pengembangan "Sistem Pencarian Terjemahan Ayat Al-Qur'an Menggunakan Teknik *Text Mining* dan *Regular Expression*". Tujuannya adalah menciptakan

sebuah alat bantu yang tidak hanya cepat dan akurat, tetapi juga intuitif sehingga mudah digunakan oleh siapa saja. Dengan menyediakan layanan informasi yang lebih baik, sistem ini diharapkan dapat membuka pintu bagi siapa pun untuk melakukan studi tematik dan mendalami pesan-pesan Al-Qur'an secara lebih personal, sesuai dengan semangat zaman digital.

Permasalahan yang dapat diidentifikasi yaitu : Pencarian ayat Al-Qur'an secara manual berdasarkan tema sangat tidak efisien, Fitur pencarian pada aplikasi digital yang ada seringkali tidak akurat dan hanya mencocokkan teks secara harfiah, Pencarian sederhana gagal membedakan kata kunci utuh dengan kata turunannya, sehingga hasilnya kurang relevan, Dibutuhkan alat bantu pencarian tematik yang lebih canggih untuk memenuhi kebutuhan pengguna di era digital.

Rumusan Masalah adalah Bagaimana membangun sebuah sistem pencarian terjemahan Al-Qur'an yang akurat dan efisien dengan menerapkan metode *text mining* dan *regular expression*?

Tujuan Penelitian adalah Membangun sistem pencarian terjemahan ayat Al-Qur'an berbasis *web* yang menerapkan teknik *text mining* dan *regular expression* untuk menghasilkan pencarian kata kunci yang akurat dan efisien bagi pengguna.

Untuk memecahkan masalah ini, saya mengambil serangkaian langkah yang logis dan terstruktur. Semuanya berawal dari data, yaitu dengan mengumpulkan dan mempersiapkan data terjemahan Al-Qur'an berbentuk file *CSV*. Setelah data siap, tantangan berikutnya adalah membuat komputer "mengerti" isi teks tersebut. Di sinilah teknik *text mining* berperan melalui proses seperti mengubah

semua huruf menjadi kecil (*case folding*), memecah kalimat menjadi kata-kata (*tokenisasi*), dan membuang kata-kata umum (*stopword removal*). Inti dari "kecerdasan" sistem ini terletak pada algoritma pencarian yang dibangun dengan *Regular Expression* (*Regex*) untuk memastikan pencocokan kata kunci yang presisi dan utuh. Teknologi ini tidak akan berguna tanpa antarmuka yang ramah, karena itu, saya membangun aplikasi *web* yang terdiri dari "otak" (*backend API* dengan *Python Flask*) dan "wajah" (*frontend* dengan *HTML*, *CSS*, dan *JavaScript*) yang memungkinkan pengguna mencari dan bahkan menyimpan hasilnya. Tahap terakhir, tentu saja, adalah pengujian menyeluruh untuk memastikan sistem ini benar-benar menjadi alat bantu yang akurat dan bermanfaat bagi siapa saja yang ingin mendalami Al-Qur'an.

2. TINJAUAN PUSTAKA

2.1. Teknologi Pencarian Ayat Al-Qur'an Digital

Perkembangan teknologi digital telah membuka akses yang lebih luas terhadap Al-Qur'an melalui berbagai platform daring. Namun, kemudahan ini belum sepenuhnya diimbangi oleh sistem pencarian yang mampu memahami konteks bahasa secara tepat. Sebagian besar aplikasi yang tersedia masih menerapkan metode pencocokan teks secara literal, yaitu mencocokkan kata kunci secara langsung tanpa mempertimbangkan struktur kata atau makna yang terkandung di dalamnya [6], [7].

Akibatnya, hasil pencarian sering kali tidak relevan dengan maksud pengguna. Misalnya, ketika seseorang mencari kata "adil", sistem bisa saja menampilkan ayat yang mengandung "pengadilan" atau "mengadili", yang maknanya berbeda secara semantik. Keterbatasan ini menunjukkan bahwa pencarian tematik terhadap ayat Al-Qur'an memerlukan pendekatan yang lebih sensitif terhadap bentuk dan makna kata.

2.2. Peran Text Mining dan Regular Expression

Text mining merupakan pendekatan dalam pengolahan bahasa alami yang berfungsi untuk mengekstraksi informasi penting dari teks tidak terstruktur. Dalam konteks Al-Qur'an, teknik ini telah banyak digunakan untuk mendukung analisis kandungan ayat, baik dalam kajian tematik, deteksi hukum bacaan *tajwid*, hingga klasifikasi wacana keagamaan [8], [9].

Proses *text mining* umumnya diawali dengan pra-pemrosesan teks, seperti penyeragaman huruf (*case folding*), pemisahan kata (*tokenisasi*), dan penghapusan kata umum yang tidak bermakna khusus (*stopword removal*). Pada sisi lain, *regular expression* (*Regex*) menawarkan cara pencarian pola kata yang sangat spesifik dan efisien. Dengan menggunakan simbol seperti `\b` untuk menandai batas kata, *Regex* memungkinkan pencarian yang benar-benar fokus pada kata kunci yang utuh, bukan bagian dari kata lain [10].

Kombinasi dari kedua pendekatan ini dapat menghasilkan sistem pencarian yang lebih akurat dan relevan secara kontekstual.

2.3. Kesenjangan Penelitian dan Kontribusi Studi Ini

Walaupun banyak penelitian telah mengeksplorasi digitalisasi dan analisis teks Al-Qur'an, sebagian besar dari studi tersebut masih berfokus pada pendekatan linguistik yang kompleks atau menggunakan model pembelajaran mesin yang membutuhkan sumber daya tinggi. Sementara itu, kebutuhan masyarakat umum lebih mengarah pada sistem yang ringan, mudah digunakan, dan tetap akurat dalam hasil pencariannya. Sayangnya, belum banyak upaya yang secara eksplisit menggabungkan *text mining* dan *regular expression* untuk membangun sistem pencarian tematik Al-Qur'an yang bersifat praktis dan humanis. Penelitian ini hadir untuk menjawab kebutuhan tersebut, dengan merancang sebuah sistem pencarian berbasis *web* yang mampu mengenali kata kunci secara utuh (*whole-word matching*) dan menampilkannya dalam antarmuka yang intuitif. Dengan demikian, sistem ini diharapkan mampu menjadi jembatan antara teknologi informasi dan kebutuhan spiritual masyarakat modern dalam memahami pesan-pesan Al-Qur'an.

Beberapa penelitian sebelumnya telah mengembangkan sistem pencarian dan layanan digital berbasis teks keagamaan, baik dengan pendekatan *natural language processing* (*NLP*) maupun sistem berbasis pola. Rizki et al. Mengembangkan layanan *chatbot* Islami menggunakan metode *word stemming* dan *pattern matching* untuk memahami konteks pertanyaan keagamaan. Sementara itu, Sulaiman et al. Mengaplikasikan *NLP* dalam *chatbot* pelayanan publik, menekankan pada interpretasi bahasa alami secara lebih luas. Di sisi lain, Mulyana et al. Merancang sistem informasi santri berbasis *web* untuk efisiensi manajemen data pendidikan Islam. Berbeda dari pendekatan tersebut, penelitian ini mengusulkan sistem pencarian ayat Al-Qur'an yang lebih ringan dan terfokus, menggunakan teknik *text mining* dan *regular expression* untuk menghasilkan pencarian yang akurat tanpa memerlukan pemrosesan semantik atau pembelajaran mesin, sehingga lebih efisien dan mudah diimplementasikan.

3. METODE PENELITIAN

3.1. Jenis dan Pendekatan Penelitian

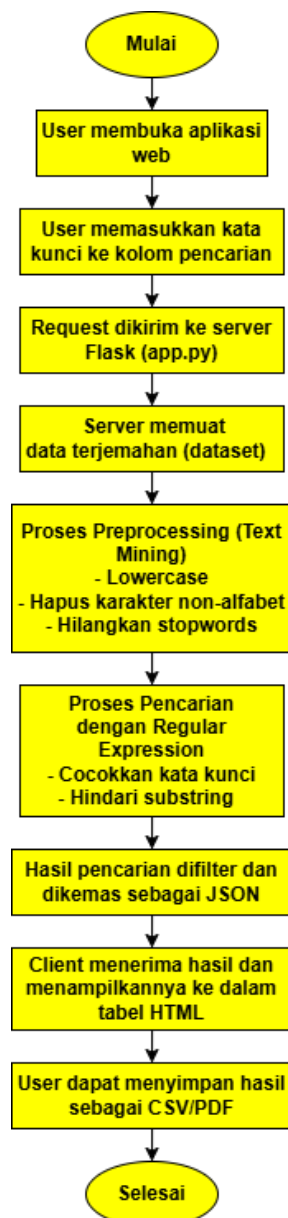
Penelitian ini merupakan penelitian terapan yang berfokus pada pengembangan sistem informasi pencarian ayat Al-Qur'an berbasis *web*. Pendekatan yang digunakan adalah deskriptif-kuantitatif, dengan tujuan untuk menjelaskan secara rinci proses perancangan sistem serta mengevaluasi kinerja sistem berdasarkan hasil pengujian fungsionalitas dan akurasi pencarian.

3.2. Data dan Sumber Dataset

Data utama yang digunakan dalam penelitian ini adalah teks terjemahan Al-Qur'an berbahasa Indonesia, yang bersumber dari situs resmi Kementerian Agama Republik Indonesia. *Dataset* ini terdiri dari 6.236 ayat terjemahan dari 114 surah, yang telah disimpan dalam format teks terstruktur. Pemilihan sumber ini dilakukan karena keotentikan dan kesesuaiannya dengan standar mushaf digital nasional.

3.3. Arsitektur Sistem dan Alur Pengembangan

Diagram dibawah ini memperlihatkan alur kerja sistem pencarian berbasis *web* yang dikembangkan.



Gambar 1. Diagram alur kerja

Sistem dikembangkan menggunakan arsitektur *client-server* berbasis *web*. Pada sisi server, digunakan *framework Python Flask* untuk menangani logika aplikasi, sementara antarmuka pengguna dibangun

menggunakan *HTML*, *CSS*, dan *JavaScript*. Sistem ini dirancang agar dapat diakses melalui *browser* tanpa memerlukan instalasi tambahan, sehingga lebih praktis bagi pengguna umum.

Alur kerja sistem dimulai dari input kata kunci oleh pengguna, yang kemudian diproses melalui mesin pencarian berbasis *regular expression*. Sistem akan mencocokkan kata kunci tersebut secara utuh (menggunakan pola `\bkeyword\b`) terhadap seluruh isi *dataset* yang telah diproses. Hasil pencarian kemudian ditampilkan dalam bentuk daftar ayat yang relevan lengkap dengan nomor surah dan ayatnya.

3.4. Tahapan Pra-pemrosesan Teks

Sebelum dilakukan pencocokan pola, data teks melalui tahap pra-pemrosesan yang meliputi:

- Case Folding*: mengubah seluruh huruf menjadi huruf kecil.
- Tokenisasi: memisahkan kalimat menjadi satuan kata.
- Penghapusan Tanda Baca dan Angka: untuk menyederhanakan struktur kalimat.
- Stopword Removal*: menghapus kata-kata umum yang tidak memiliki nilai semantik khusus.

Proses ini bertujuan untuk meningkatkan efisiensi pencarian dan meminimalkan kesalahan pencocokan.

3.5. Algoritma Pencarian Berbasis Regular Expression

Pencarian dilakukan menggunakan pola *regular expression* `\bkeyword\b`, yang memungkinkan sistem mengenali kata kunci secara utuh dengan menghindari pencocokan sebagian kata (*substring*). Teknik ini terbukti sederhana namun efektif untuk mengatasi kesalahan pencarian yang umum terjadi pada sistem konvensional. Proses pencarian dilakukan secara iteratif terhadap seluruh *dataset*, dengan hasil ditampilkan secara *real-time*.

3.6. Pengujian Sistem

Pengujian dilakukan dengan dua metode utama:

- Pengujian fungsional (*black-box testing*) untuk memastikan bahwa setiap fitur berjalan sesuai kebutuhan, seperti *input*, *output*, dan navigasi halaman.
- Uji komparatif sederhana terhadap sistem pencarian literal (tanpa *regex*) guna melihat perbedaan hasil pencarian. Penilaian dilakukan berdasarkan relevansi hasil dan keutuhan kata yang dicocokkan.

Hasil pengujian menunjukkan bahwa sistem mampu menyaring hasil pencarian secara lebih akurat dan menghindari pencocokan yang ambigu atau tidak sesuai konteks.

4. HASIL DAN PEMBAHASAN

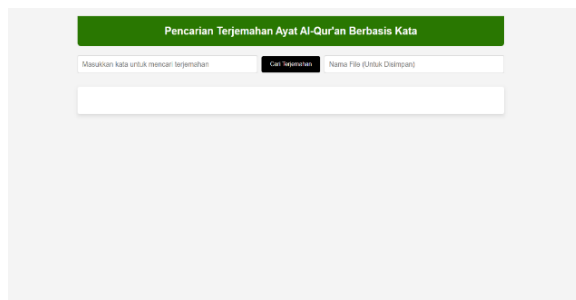
4.1. Implementasi Sistem Pencarian

Penelitian ini menghasilkan sebuah aplikasi *web* yang berfungsi sebagai sistem pencarian ayat Al-

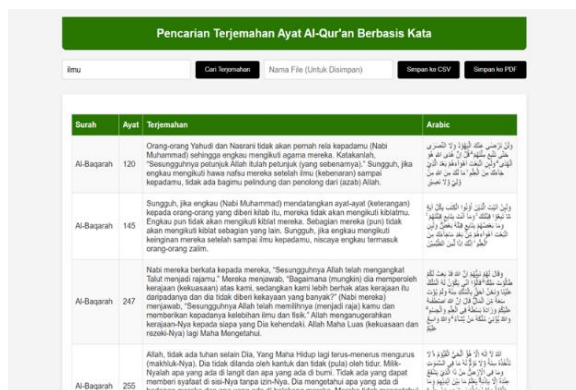
Qur'an berbasis kata kunci dalam terjemahan Bahasa Indonesia. Aplikasi ini dikembangkan menggunakan *framework Python Flask* pada sisi server, dan memanfaatkan teknologi *HTML, CSS*, serta *JavaScript* untuk membangun antarmuka pengguna. Pengguna cukup memasukkan kata kunci yang ingin dicari, dan sistem akan menampilkan daftar ayat yang mengandung kata tersebut secara utuh, lengkap dengan nama surah dan nomor ayat.

Antarmuka sistem dirancang dengan pendekatan sederhana dan responsif, sehingga dapat digunakan secara nyaman baik melalui perangkat desktop maupun seluler. Keunggulan lainnya adalah sistem ini tidak memerlukan instalasi atau proses pendaftaran akun, sehingga pengguna dari berbagai kalangan dapat langsung mengakses dan memanfaatkannya dengan mudah.

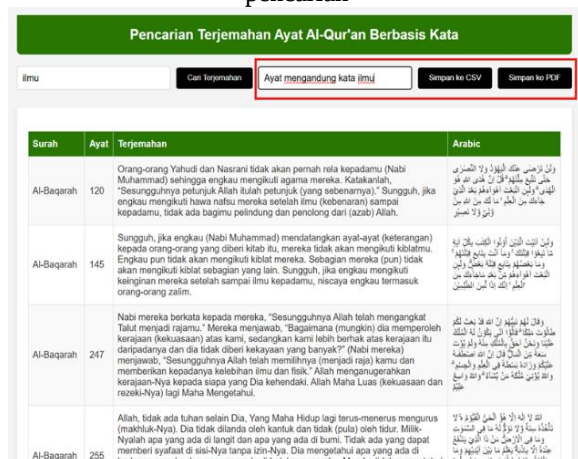
Gambar dibawah ini menampilkan tampilan antarmuka sistem saat pengguna melakukan pencarian kata kunci.



Gambar 2. Tampilan awal website



Gambar 3. Tampilan saat website menampilkan hasil pencarian



Gambar 4. Tampilan fitur penyimpanan ke csv atau pdf

4.2. Hasil Pencarian dan Ketepatan Pola Pencocokan

Sistem ini mengandalkan pola pencocokan kata secara utuh atau *whole-word matching*, yang diwujudkan melalui penggunaan regular expression dengan pola `\bkeyword\b`. Pendekatan ini memungkinkan sistem untuk membedakan antara kata dasar dan bentuk turunannya. Sebagai contoh, ketika pengguna mencari kata "iman", sistem tidak akan menampilkan ayat dengan kata "beriman", "keimanan", atau bentuk turunan lainnya. Hal ini memastikan bahwa hasil pencarian benar-benar relevan dengan kata kunci yang dimaksud, sehingga dapat menghindari kesalahan interpretasi yang mungkin muncul akibat pencocokan parsial.

Dalam pengujian sederhana, sistem menunjukkan peningkatan relevansi hasil pencarian dibandingkan dengan metode pencarian literal. Jika pencarian biasa menghasilkan banyak ayat yang tidak sepenuhnya sesuai dengan makna kata kunci, sistem ini berhasil menyaring hasil secara lebih tepat, sehingga dapat lebih mendukung pengguna dalam memahami kandungan ayat secara tematik.

Berikut dibawah ini adalah tabel Perbandingan hasil dan ketepatan pola pencocokan.

Tabel 1. Perbandingan hasil dan ketepatan pola pencocokan

No.	Kata Kunci Uji	Hasil Sistem (Penerapan RegEx)	Hasil Pencarian Konvensional (Ctrl+F)	Analisis Akurasi	Status
1	"aman"	Hanya menampilkan ayat dengan kata "aman" yang utuh.	Menampilkan ayat dengan kata "aman", "keamanan", dll.	Sistem terbukti lebih akurat karena berhasil menyaring hasil yang tidak relevan (<i>false positive</i>).	Berhasil
2	"ilmu"	Hanya menampilkan ayat dengan kata "ilmu" yang utuh.	Menampilkan ayat dengan kata "ilmu", "ilmuwan", dll.	Hasil pencarian sistem lebih presisi dan sesuai dengan intensi pengguna.	Berhasil
3	"adil"	Hanya menampilkan ayat dengan kata "adil" yang utuh.	Menampilkan ayat dengan kata "pengadilan", dll.	Metode <i>whole-word matching</i> terbukti efektif mengeliminasi hasil yang ambigu.	Berhasil

4.3. Pengujian Fungsionalitas Sistem

Untuk memastikan keandalan sistem, dilakukan pengujian dengan pendekatan *black-box* yang berfokus pada fungsi-fungsi utama aplikasi. Beberapa skenario pengujian meliputi: proses *input* kata kunci, pencocokan hasil terhadap *dataset*, tampilan ayat yang ditemukan, serta penanganan terhadap kondisi kosong atau tidak ditemukan. Seluruh skenario pengujian berhasil dijalankan dengan baik, menunjukkan bahwa

sistem bekerja secara stabil dan sesuai dengan kebutuhan pengguna.

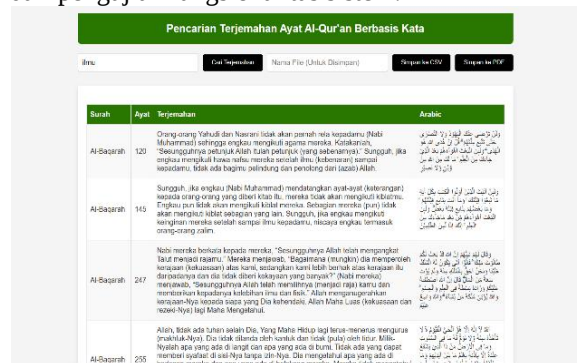
Respons waktu pencarian tetap cepat meskipun sistem bekerja dengan *dataset* yang cukup besar, mencakup ribuan ayat Al-Qur'an. Selain itu, sistem juga memberikan pesan umpan balik yang jelas saat tidak ditemukan hasil, sehingga tetap memberikan pengalaman pengguna yang baik.

Tabel dibawah ini menunjukkan skenario pengujian fungsionalitas sistem.

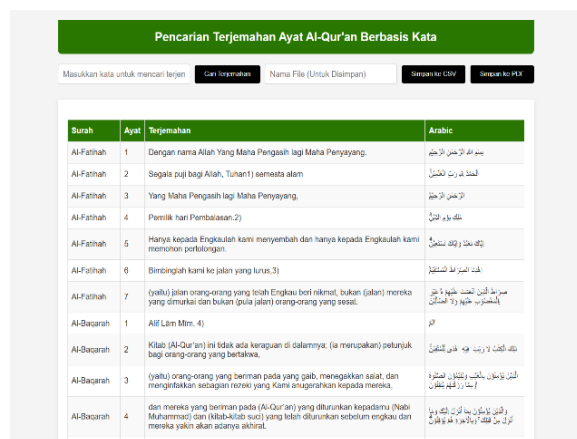
Tabel 2. Skenario pengujian fungsionalitas sistem

No	Skenario Uji	Input	Hasil yang Diharapkan	Hasil Aktual	Status
1	Klik tombol setelah input kata	ilmu	Hasil pencarian muncul dalam tabel	Sesuai	Berhasil
2	Input kosong	(kosong)	Tidak crash, tampilkan seluruh ayat yang ada di dataset Al-Qur'an	Sesuai	Berhasil
3	Input simbol/angka	123@	Sistem tetap stabil, tidak error	Sesuai	Berhasil

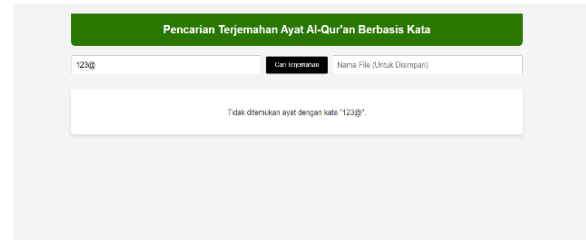
Adapun Gambar dibawah ini menunjukkan hasil dari pengujian fungsionalitas sistem.



Gambar 5. Pengujian klik tombol setelah *input* kata “ilmu”



Gambar 6. Pengujian *input* kosong



Gambar 7. Pengujian simbol/angka

Sebagai pelengkap dari fungsi pencarian utama, sistem ini juga menyediakan fitur ekspor hasil pencarian ke dalam format CSV dan PDF yang bertujuan memudahkan pengguna dalam menyimpan, membagikan, atau mendokumentasikan hasil pencarian untuk keperluan studi maupun arsip pribadi. Berdasarkan pengujian, kedua format berjalan dengan baik: file CSV yang dihasilkan memiliki struktur data yang konsisten dan dapat dibuka di berbagai aplikasi *spreadsheet*, sementara file PDF tersaji dengan tampilan yang rapi dan mudah dibaca. Seluruh data yang ditampilkan pada antarmuka sistem terekspor secara utuh tanpa kehilangan informasi atau gangguan karakter, baik pada pencarian dengan sedikit maupun banyak hasil. Fitur ini menambah nilai guna sistem dengan memberikan fleksibilitas tambahan bagi pengguna yang membutuhkan salinan hasil dalam bentuk yang dapat diakses secara *offline* atau dicetak.

Tabel dibawah ini menunjukkan skenario pengujian fitur ekspor hasil ke CSV atau PDF.

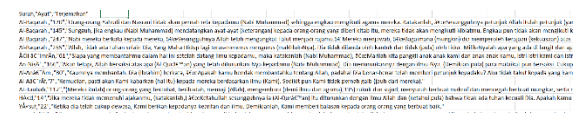
Tabel 3. Skenario pengujian fitur ekspor hasil

No	Deskripsi Uji	Format	Hasil yang Diharapkan	Hasil Aktual	Status
1	Ekspor ke CSV setelah pencarian kata	CSV	File CSV berisi semua hasil pencarian	Sesuai	Berhasil
2	Ekspor ke PDF setelah pencarian kata	PDF	File PDF memiliki struktur tabel dan teks yang rapi	Sesuai	Berhasil

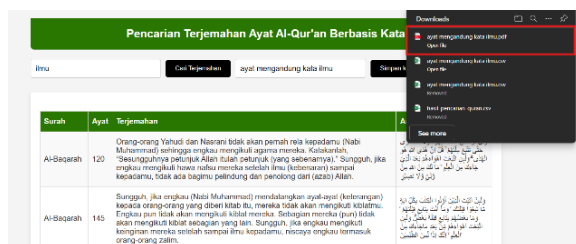
Adapun Gambar dibawah ini menunjukkan hasil dari pengujian fitur ekspor hasil.



Gambar 8. Pengujian fitur simpan CSV



Gambar 9. Hasil dari ekspor CSV



Gambar 10. Pengujian fitur simpan PDF

Hasil Pencarian Ayat Al-Qur'an

Surah	Ayat	Terjemahan
Al-Baqarah	120	Orang-orang Yahudi dan Nasrani tidak akan pernah ada kepadamu (Nabi Muhammad) sehingga engkau mengikuti agamamu. Katakanlah, "Seungguhnya petunjuk Allah itu satu (yang sebenarnya)." Sungguh, jika engkau mengikuti hawa nafsu mereka setelah ilmu (keberatan) sampai kepadamu, tidak ada bagimu petunjuk dan penolong dari (sisi) Allah.
Al-Baqarah	145	Sungguh, jika engkau (Nabi Muhammad) mendatangkan ayat-ayat (keberangan) kepada orang-orang yang diberi kitab itu, mereka tidak akan mengikutimu. Engkau pun tidak akan mengikuti kitab mereka. Sebagian mereka (pun) tidak akan mengikuti kitab sebagian yang lain. Sungguh, jika engkau mengikuti keinginan mereka setelah ilmu kepadamu, niscaya engkau termasuk orang-orang zalim.
Al-Baqarah	247	Nabi mereka berkata kepada mereka, "Seungguhnya Allah telah mengangkat Thalut menjadi rajanya." Mereka menjawab, "Bagaimana (mungkin) dia memperoleh kerajaan (kekuasaan) atas kami, sedangkan kami lebih baik dari dia?" Thalut menjawab, "Dia telah memilihnya (menjadi raja) karena dia memberikan kepadamu kelebihan ilmu dan fisik." Allah mengantarannya kepada mereka sebagai tanda yang Dia kehendaki. Allah Maha Luas (kekuasaan dan rezeki-Nya) bagi Maha Mengetahui.
Al-Baqarah	255	Allah, tidak ada tuhan selain Dia. Yang Maha Hidup lagi terus-menerus mengurus (malaikat-Nya). Dia tidak dilanda oleh kantuk dan tidak (pula) oleh tidur. Milik-Nyalah apa yang ada di langit dan apa yang ada di bumi. Tidak ada yang dapat memberi syafaat di sisi-Nya tanpa izin-Nya. Dia mengetahui apa yang ada di hadapan mereka dan apa yang ada di belakang mereka. Mereka tidak mengetahui sesuatu apa pun dari ilmu-Nya, kecuali apa yang Dia kehendaki. Kuno-Nya (ilmu dan kekuasaan-Nya) meliputi langit dan bumi. Dia tidak merasa terburu-buru mendaliranya. Dia-lah yang Maha Tinggi lagi Maha Agung.

Gambar 11. Hasil dari ekspor PDF

4.4. Refleksi Hasil dan Potensi Pengembangan

Hasil implementasi menunjukkan bahwa sistem ini dapat menjadi solusi yang efektif untuk kebutuhan pencarian ayat tematik dalam Al-Qur'an. Dengan pendekatan *regular expression*, pencarian dapat dilakukan dengan presisi tinggi tanpa memerlukan algoritma pembelajaran mesin yang kompleks. Hal ini membuat sistem lebih ringan, cepat, dan mudah diakses oleh pengguna dari berbagai latar belakang.

Namun demikian, sistem ini masih memiliki keterbatasan, khususnya dalam mengenali variasi semantik seperti sinonim atau makna tersirat. Ke depan, pengembangan sistem dapat diarahkan untuk mengintegrasikan teknologi pendukung seperti kamus sinonim atau teknik *stemming*, guna meningkatkan fleksibilitas pencarian. Dengan demikian, sistem ini

tidak hanya akan akurat dari sisi teknis, tetapi juga lebih kontekstual dari sisi makna.

5. KESIMPULAN DAN SARAN

Penelitian ini menghasilkan sebuah sistem pencarian ayat Al-Qur'an berbasis web yang dirancang untuk meningkatkan akurasi pencarian melalui penerapan teknik *text mining* dan *regular expression* dengan pendekatan *whole-word matching*. Sistem ini mampu menyaring hasil pencarian secara lebih tepat, menghindari pencocokan yang ambigu, serta memberikan pengalaman penggunaan yang sederhana dan ramah bagi pengguna dari berbagai latar belakang. Secara keseluruhan, sistem menunjukkan performa fungsional yang baik dan dapat menjadi solusi ringan namun efektif dalam mendukung pencarian tematik terhadap ayat-ayat Al-Qur'an terjemahan. Meski demikian, sistem masih memiliki keterbatasan dalam mengenali variasi bentuk kata dan makna semantik yang lebih dalam. Oleh karena itu, pengembangan di masa mendatang dapat diarahkan pada integrasi teknik *stemming*, pemetaan sinonim, atau pendekatan *NLP* berbasis konteks agar sistem menjadi lebih adaptif dan inklusif terhadap ragam ekspresi bahasa yang digunakan oleh pengguna.

DAFTAR PUSTAKA

- [1] Binti Khabibatur R.A., "Digitalisasi Al-Qur'an: Transformasi dan Sakralitas Al-Qur'an 504," 2025.
- [2] D. I. Mulyana and M. A. I. Rowis, "Optimization of Text Mining Detection of Tajweed Reading Laws Using the Yolov8 Method on the Qur'an," *QALAMUNA: Jurnal Pendidikan, Sosial, dan Agama*, vol. 14, no. 2, pp. 1089–1110, Dec. 2022, doi: 10.37680/qalamuna.v14i2.3866.
- [3] J. Khatib Sulaiman, M. Abid Nadzif, and U. Stikubank Semarang, "Penggunaan Teknologi Natural Language Processing dalam Sistem Chatbot Untuk Peningkatan Layanan Informasi Administrasi Publik," *Indonesian Journal of Computer Science Attribution*, vol. 13, no. 1, pp. 2024–1227, 2024.
- [4] S. A. van Laar, K. B. Gombert-Handoko, H. J. Guchelaar, and J. Zwaveling, "An Electronic Health Record Text Mining Tool to Collect Real-World Drug Treatment Outcomes: A Validation Study in Patients With Metastatic Renal Cell Carcinoma," *Clin Pharmacol Ther*, vol. 108, no. 3, pp. 644–652, Sep. 2020, doi: 10.1002/cpt.1966.
- [5] M. Rizki, A. Fitriyansyah, and M. Narji, "APLIKASI CHATBOT SEBAGAI LAYANAN LIVE CHAT UNTUK PENERIMAAN MAHASISWA BARU MENGGUNAKAN METODE WORD STEMMING DENGAN REGULAR EXPRESSION PATTERN MATCHING," 2023.
- [6] T. Margoni and M. Kretschmer, "A Deeper Look into the EU Text and Data Mining Exceptions:

- Harmonisation, Data Ownership, and the Future of Technology,” GRUR International*, vol. 71, no. 8, pp. 685–701, Aug. 2022, doi: 10.1093/grurint/ikac054.
- [7] Z. Liu and Y. Lin Maosong Sun Editors, “*Representation Learning for Natural Language Processing Second Edition.*”
- [8] S. Choo and W. Kim, “A study on the evaluation of tokenizer performance in natural language processing,” *Applied Artificial Intelligence*, vol. 37, no. 1, 2023, doi: 10.1080/08839514.2023.2175112.
- [9] E. Uzun, “A regular expression generator based on CSS selectors for efficient extraction from HTML pages,” *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 28, no. 6, pp. 3389–3401, Aug. 2020, doi: 10.3906/ELK-2004-67.
- [10] K. G. Srinivasa, G. M. Siddesh, and S. R. Manisekhar Editors, “*Statistical Modelling and Machine Learning Principles for Bioinformatics Techniques, Tools, and Applications.*” [Online]. Available: <http://www.springer.com/series/16171>