

PENERAPAN METODE RULE BASED PADA QUESTION ANSWERING SYSTEM TERJEMAH AL-QURAN BERTEMA SAINS

Ade Vita Aulia, Badrus Zaman, dan Rimuljo Hendradi

Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Airlangga
Surabaya, Indonesia

e-mail: badruszaman@fst.unair.ac.id

ABSTRAK

Sekitar seperdelapan isi dari Al Quran sebagai kitab suci umat Islam, terdiri dari ayat-ayat tentang dasar-dasar ilmu pengetahuan (sains). Meskipun ayat-ayat al-Qur'an tersebut telah terklasifikasikan menurut temanya, namun diperlukan waktu untuk mencari kembali ayat-ayat sesuai dengan tema sains pada permasalahan tertentu. Untuk memudahkan pencarian tersebut dapat digunakan Question Answering System (QAS), sehingga didapatkan jawaban yang relevan dengan pertanyaan. Metode yang digunakan dalam QAS terdiri dari tiga tahap, yaitu question analysis, passage retrieval, dan answer extraction. Tahap question analysis dilakukan untuk melakukan analisis pertanyaan. Tahap passage retrieval dilakukan untuk mengambil dokumen-dokumen ayat terjemahan menggunakan TF-IDF dan cosine similarity. Sedangkan tahap answer extraction dilakukan untuk ekstraksi hasil jawaban dengan menggunakan metode word-match dan rule-based. Data yang digunakan untuk ujicoba sebanyak 387 ayat yang terbagi dalam 10 sub-tema. Berdasarkan hasil evaluasi, sistem menunjukkan hasil kinerja yang baik. Hal ini terlihat hasil evaluasi sistem dengan menggunakan 20 pertanyaan didapatkan hasil precision, recall dan f-measures berturut-turut sebesar 0.75, 0.72 dan 0.74.

Kata Kunci: *al-Qur'an, information retrieval, question answering system (QAS), rule-based*

ABSTRACT

The Qur'an is the holy book of Muslims where about one-eighth are verses hint at the basics of science. Although the verses of the Qur'an have been classified according to the theme, it takes time to search for verses according to the theme of science on certain problems. To facilitate the search, a Question Answering System (QAS) can be used, so that answers are obtained that are relevant to the question. The method used in QAS consists of three stages, namely question analysis, passage retrieval, and answer extraction. The question analysis stage is done to analyze the question. The passage retrieval stage is done to retrieve translated verse documents using TF-IDF and cosine similarity. While the answer extraction stage is done to extract the answer results using word-match and rule-based methods. The data used for testing is 387 verses divided into 10 sub-themes. Based on the evaluation results, the system shows excellent performance results. This can be seen from the system evaluation results using 20 questions obtained precision, recall and f-measures of 0.75, 0.72 and 0.74 respectively.

Keyword: *al-Qur'an, information retrieval, question answering system (QAS), rule-based*

I. PENDAHULUAN

Al-Qur'an, sebagai kitab suci umat Islam, memiliki signifikansi yang mendalam dalam berbagai aspek kehidupan, termasuk ilmu pengetahuan. Dari total 6.236 ayat yang terdapat dalam Al-Qur'an, sekitar seperdelapan atau sekitar 750 ayat merupakan ayat-ayat yang berkaitan dengan berbagai aspek sains [1,2]. Untuk memudahkan akses dan penelusuran terhadap konten ilmiah yang terdapat dalam Al-Qur'an, beberapa metode pengelompokan ayat berdasarkan tema tertentu telah dikembangkan, dan indeks klasifikasi ayat Al-Qur'an menjadi salah satunya [1]. Meskipun indeks tersebut memberikan struktur yang terorganisir untuk penelusuran ayat-ayat yang sesuai dengan topik tertentu, namun proses mencari kembali ayat-ayat dalam klaster atau tema yang sama

masih memerlukan waktu dan usaha yang cukup besar. Oleh karena itu, terdapat kebutuhan untuk mengembangkan sistem yang dapat mengotomatisasi proses pencarian dan analisis makna dari ayat-ayat Al-Qur'an yang berkaitan dengan sains. Pengembangan sistem semacam ini dapat melibatkan teknologi-teknologi baru seperti pemrosesan bahasa alami (NLP) dan kecerdasan buatan (AI) untuk melakukan analisis konten Al-Qur'an secara otomatis menggunakan Question Answering System (QAS).

QAS adalah sistem komputer yang dapat menjawab pertanyaan dari pengguna dengan menggunakan kata kunci berupa pertanyaan dalam bahasa alami. Sistem ini memberikan jawaban spesifik dan jelas sesuai dengan kebutuhan pengguna [3,4]. Tujuan dari QAS sebenarnya serupa dengan mesin pencari, yaitu mencari informasi dalam dokumen dengan kata kunci. Namun, perbedaannya terletak pada cara penyajiannya dimana

QAS memberikan jawaban langsung berdasarkan pertanyaan pengguna, sementara mesin pencari mengembalikan daftar dokumen yang relevan dengan kata kunci [3-5].

Implementasi QAS untuk temu kembali telah dilakukan dalam beberapa penelitian. QAS dengan metode berbasis aturan sudah diterapkan pada terjemahan al-Qur'an berbahasa Indonesia pada surat al-Baqarah, dengan fokus pada pertanyaan dengan kata tanya "siapa", "apa", "kapan", "mengapa", dan "mana" [6]. Hasil Evaluasi menunjukkan sistem tersebut memiliki akurasi jawaban yang memuaskan. Pada penelitian lainnya, telah dikembangkan QAS Bahasa Indonesia untuk pertanyaan yang non-factoid, seperti "siapa", "apa", "kapan", "mengapa", dan "mana" [5]. Hasilnya menunjukkan nilai rata-rata urutan teratas dari jawaban yang benar cukup tinggi. Sedangkan QAS non-factoid lainnya pada terjemahan Bahasa Indonesia surat al-Baqarah, dengan menggunakan pola kalimat dari hasil analisis dokumen menunjukkan kinerja sistem yang lebih baik daripada sistem yang tanpa analisis dokumen terlebih dahulu atau dengan kata lain hanya menggunakan pola kalimat [6].

Berdasarkan penelitian terdahulu, penelitian ini berfokus pada implementasi QAS pada terjemahan Al-Qur'an yang berfokus pada tema sains. Metode yang akan digunakan adalah rule-based, di mana sejumlah aturan akan digunakan untuk mencari kalimat jawaban dalam dokumen terjemahan ayat yang tersimpan. Pendekatan rule-based telah terbukti efektif dalam penelitian sebelumnya, seperti pada penerapan metode rule-based pada terjemahan surat Al-Baqarah dengan tipe pertanyaan berupa kata tanya seperti "siapa", "apa", "kapan", "mengapa", dan "mana" [5].

Aturan yang digunakan untuk mengidentifikasi kalimat-kalimat dalam dokumen terjemahan ayat yang berpotensi menjadi jawaban, berdasarkan tipe pertanyaan yang diajukan. Pengelompokan tipe jawaban berdasarkan tipe pertanyaan akan dilakukan dengan menggunakan pola kalimat dalam bahasa Indonesia. Misalnya, jawaban dari pertanyaan "apa" akan diidentifikasi dengan kata-kata seperti "adalah" atau "yaitu", sementara jawaban dari pertanyaan "dimana" akan diidentifikasi dengan kata-kata seperti "dalam", "dari", atau "pada". Pembobotan jawaban akan dilakukan berdasarkan kemiripan antara kata-kunci pertanyaan dan kalimat dokumen terjemahan ayat, serta bobot dari masing-masing aturan berdasarkan tipe pertanyaan. Kalimat dengan skor tertinggi akan dipilih sebagai jawaban, dan hasilnya akan dievaluasi melalui penilaian manusia atau pakar.

Dengan menggunakan pendekatan ini diharapkan QAS dapat memberikan jawaban yang lebih akurat dan relevan terhadap pertanyaan-pertanyaan terkait sains dalam Al-Qur'an. Evaluasi hasil jawaban dari sistem

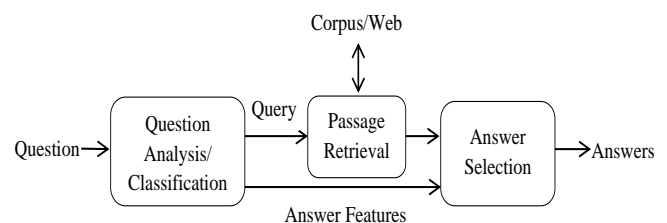
QA ini akan memberikan wawasan yang berharga tentang efektivitas metode rule-based dalam konteks penelitian ini.

II. TINJAUAN PUSTAKA

Dari sudut pandang ilmu al-Quran, ayat-ayat al-Qur'an dapat dibagi menjadi dua jenis, yaitu ayat kauniyah dan qauliyah. Ayat kauniyah merujuk pada tanda-tanda Allah yang ada di alam semesta, sementara ayat qauliyah adalah firman Allah SWT yang tertulis dalam kitab suci-Nya. Dari kedua jenis ayat ini, berkembang berbagai ilmu yang dipelajari manusia, seperti ilmu kedokteran, astronomi, kimia, matematika, teknik, tasawuf, fiqih, nahwu, dan lain sebagainya [7]. Al-Qur'an memiliki banyak mukjizat, di antaranya kandungan ayat-ayat ilmu pengetahuan, yang jumlahnya hingga 1000 ayat. Analisis ilmiah menunjukkan bahwa ayat-ayat ini mengandung informasi yang faktual dan konsisten dengan temuan ilmiah terbaru[8].

Question Answering (QA) system adalah sebuah sistem cerdas yang dapat memberikan jawaban benar dari suatu pertanyaan yang diajukan oleh *user* di mana akurasi jawaban dipengaruhi oleh *rule* atau fakta-fakta yang tersimpan pada *knowledge base* [9]. Berdasarkan arsitektur umum, sistem QA terbagi atas tiga komponen utama, yaitu analisis/klasifikasi pertanyaan, pencarian teks, dan pemilihan jawaban, seperti yang digambarkan dalam Gambar 1. [5].

Rule-Based merupakan cara sederhana dalam menentukan kategori pertanyaan masukan dengan menggunakan suatu aturan yang dibuat secara manual oleh pakar di bidangnya. Salah satu implementasi rule based pada QA *system* adalah Quarc (*Question Answering for Reading Comprehension*) [9,10].



Gambar 1. Arsitektur *Question Answering System* (Croft et al., 2010)

Anggraeny (2007) melakukan penelitian tentang QA *system* pada terjemahan al-Qur'an surat al-Baqarah berbahasa Indonesia dengan menerapkan metode *rule based* yang sebelumnya telah digunakan oleh Riloff & Thelen (2000). Penelitian ini bertujuan untuk melakukan modifikasi algoritma rule-based untuk sistem pencocokan pola kalimat. Algoritma rule-based ini disusun berdasarkan analisis pola kalimat dalam

terjemahan surat al-Baqarah dan pola kalimat kueri bahasa alami. Algoritma rule yang disusun ini merupakan modifikasi dari algoritma rule based yang sudah ada, dengan mempertimbangkan kaidah tata bahasa Indonesia.

Untuk mengevaluasi kinerja sistem temu kembali informasi, digunakan tiga metrik utama: *precision* (persamaan 1), *recall* (persamaan 2), dan *f-measure* (persamaan 3). Metrik-metrik ini dihitung berdasarkan perhitungan matematis yang melibatkan elemen-elemen relevan dan tidak relevan dalam hasil pencarian.

$$Precision = \frac{\text{jumlah kalimat relevan yang dihasilkan sistem}}{\text{jumlah kalimat yang dihasilkan sistem}} \quad [1]$$

$$Recall = \frac{\text{jumlah kalimat relevan yang dihasilkan sistem}}{\text{jumlah kalimat relevan}} \quad [2]$$

$$f - \text{measures} = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad [3]$$

III. METODE PENELITIAN

Penelitian ini secara umum dilakukan dengan empat tahap, yaitu pengumpulan data, pengolahan dokumen, implementasi *question answering system* dengan metode rule-based, dan evaluasi sistem. Masing-masing tahap dijelaskan sebagai berikut.

A. Pengumpulan Data

Data yang digunakan dalam penelitian ini berupa dokumen yang terdiri dari kumpulan terjemahan ayat-ayat al-Qur'an dengan tema sains.

B. Pengolahan Dokumen

Dokumen berupa data terjemah ayat dilakukan pre-processing data yang meliputi *tokenizing*, *stopword removal* dan *stemming*. Pengolahan dokumen ini diperlukan untuk menormalisasi data teks sebelum masuk ke tahap selanjutnya. Hasil normalisasi data terjemah ayat disimpan pada file yang sama dengan dokumen aslinya.

1. *Tokenizing*: proses pemisahan rangkaian kata pada kalimat, paragraf atau dokumen menjadi *token* atau potongan kata tunggal yang berdiri sendiri. Dalam proses *tokenizing*, karakter dan symbol selain a-z dihilangkan. Karakter-karakter tertentu juga dihilangkan pada tahapan ini, seperti tanda baca dan diubah menjadi bentuk huruf kecil (*lower case*). Secara umum pengambilan token pada suatu kalimat dilakukan dengan cara memisahkan semua kata berdasarkan adanya pemisah (*delimiter*) *whitespace* seperti spasi, tab, dan *new line* (enter). Namun ada juga karakter seperti tanda petik ("), titik (.), titik dua (:), semikolon (;), atau lainnya yang berperan sebagai pemisah kata. Kemudian tanda *hyphen* (-) yang biasanya digunakan untuk menggabungkan

dua token yang berbeda, rentang nilai, kata berulang, dan lain sebagainya. Sedangkan, karakter *slash* (/) untuk menyatakan "dan atau", pemisah file atau direktori atau *url*.

2. *Stopword removal*: kata hasil *tokenizing* dibandingkan dengan stopwords, jika kata tersebut ada di dalam *stopwords* maka kata tersebut dihapus. *Stopwords* yang digunakan sesuai dengan *stopwords* Anggraeny (2007).
3. *Stemming*: kata hasil *stopword removal* akan dilakukan proses penghapusan imbuhan berupa awalan, akhiran atau gabungan awalan dan akhiran untuk mendapatkan kata dasar menggunakan *library* Sastrawi berbasis Python.

C. Implementasi QAS dengan Metode Rule Based

Arsitektur *Question Answering System* menggunakan metode berbasis aturan (*rule based*) seperti ditunjukkan pada Gambar 2.

Di dalam QA system terdapat tiga tahapan utama yang bekerja untuk menemukan jawaban berdasarkan kata kunci pertanyaan yang dimasukkan oleh *user* yaitu *question analysis*, *passage retrieval*, dan *answer extraction*. Masing-masing tahap dijelaskan sebagai berikut.

1. Question Analysis

Kata kunci pertanyaan yang telah dimasukkan oleh *user* diambil oleh sistem yang selanjutnya akan melalui tahap *tokenizing*, *stopword removal* dan *stemming*. Sistem juga mengambil informasi tentang jenis kata tanya yang digunakan pada kata kunci pertanyaan dan menyimpannya sementara didalam sistem.

2. Passage Retrieval

Kata kunci yang telah dinormalisasi pada tahap *question analysis* selanjutnya akan dilakukan pencocokan kesamaan setiap kata pada kata kunci terhadap data terjemah ayat yang telah disimpan pada tahap pengolahan dokumen. Hasil pencocokan kata tersebut mengembalikan dokumen-dokumen yang relevan berupa data terjemah ayat yang memiliki minimal satu kata yang sama dengan kata kunci. Pengambilan dokumen-dokumen terjemah ayat relevan dengan kata kunci tersebut melalui proses pembobotan dokumen terjemah ayat dengan TF-IDF dan proses mencari kemiripan dokumen terjemah ayat dengan kata kunci menggunakan *cosine similarity*.

3. Answer Extraction

Dokumen-dokumen terjemah ayat relevan berupa kalimat yang telah terambil pada tahap *passage retrieval*, akan dicocokkan kembali dengan kata kunci pertanyaan menggunakan algoritma *wordMatch* dan *rule based*. Proses pemberian nilai menggunakan algoritma *WordMatch* dilakukan saat

mencari kesamaan token setiap kalimat dalam dokumen terjemah ayat dengan token pada kata kunci pertanyaan, sedangkan *rule based* digunakan untuk memberikan *scoring* pada dokumen terjemah ayat yang dicocokkan berdasarkan aturan yang telah dibuat. Kemudian jawaban yang ditentukan oleh sistem berdasarkan dokumen terjemah dengan tingkat relevansi tertinggi.

D. Evaluasi Sistem

Evaluasi sistem dilakukan dengan membandingkan jawaban yang dihasilkan oleh sistem dengan jawaban yang relevan (*gold standard*). Jawaban yang relevan ditentukan berdasarkan pendapat 2 orang pakar, dimana jawaban dianggap relevan jika dipilih oleh salah satu pakar.

Jawaban sistem akan dianggap benar jika jawaban merujuk pada ayat dan surat yang sama dengan jawaban pakar. Kemudian, dari hasil perbandingan tersebut, maka akan didapatkan nilai *precision*, *recall* dan *f-measures*.

IV. HASIL

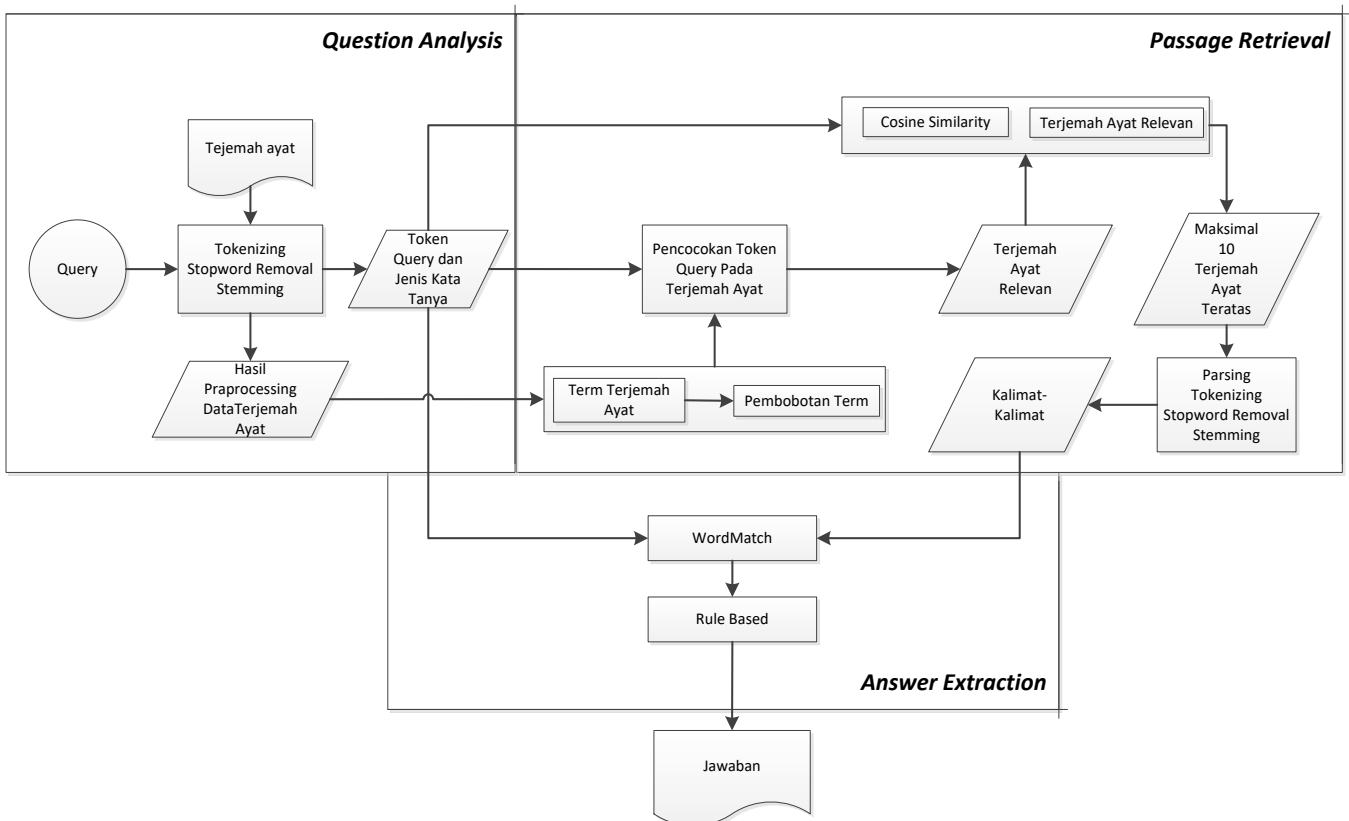
A. Pengumpulan Data

Penelitian ini menggunakan data terjemahan al-Qur'an dengan tema sains sebanyak 387 ayat yang dikelompokkan ke dalam 58 sub-tema. Data tersebut berasal dari 2 sumber, yaitu terjemah al-Qur'an Kementerian Agama Republik Indonesia dan penelitian

dari Purwaningrum (2015) [7].

1. Terjemah al-Qur'an tersebut diterbitkan pada tahun 2010 oleh Syaamil Quran. Pengumpulan terjemahan ayat-ayat al-Qur'an yang bertema sains diambil berdasarkan indeks klasifikasi ayat al-Qur'an yang terdapat pada bab ilmu dan cabang-cabangnya di mana pada bab tersebut masih terbagi ke dalam beberapa sub-bab. Pengambilan ayat-ayat al-Qur'an yang memuat tema sains dilakukan pada sub-bab al-falak, bintang-bintang, kalender, perkapalan, seni-seni, bukti-bukti ilmiah dan isyarat terhadap kejadian-kejadian penting yang telah dibuktikan oleh penemuan ilmiah modern.
2. Penelitian Purwaningrum (2015) berupa ayat-ayat dikelompokkan berdasarkan disiplin suatu ilmu yang terbagi menjadi Ilmu Pengetahuan Alam (IPA), Fisika, Kimia, Matematika, Bahasa, Teknik Informatika, dan Teknik Arsitektur.

Satu terjemahan ayat al-Qur'an direpresentasikan dengan satu dokumen, kecuali pada ayat yang memiliki kelanjutan pada ayat berikutnya, maka ayat tersebut akan digabung menjadi satu dokumen. Dalam satu dokumen minimal berisi satu kalimat, dimana kalimat tersebut terdiri dari subjek dan predikat. Setiap kalimat diakhiri dengan tanda baca titik (.), atau ayat yang kalimatnya diakhiri dengan tanda tanya (?) atau tanda seru (!), namun tidak berhubungan dengan ayat selanjutnya. Pada penelitian ini ayat al-Quran bertema sains yang digunakan terdapat 387 ayat. Ayat-ayat ini



Gambar 2. Arsitektur *Question Answering System (QAS)* menggunakan *rule-based*

dikelompokkan berdasarkan 10 sub tema. Jumlah ayat pada masing-masing subtema ditunjukkan pada Tabel 1.

TABEL I
PENGELOMPOKAN AYAT AL-QURAN YANG BERTEMA SAINS

No	Sub Tema	Jumlah Ayat
1	Fisika	78
2	Kimia	16
3	Matematika dan Informatika	17
4	Seni dan Bahasa	13
5	Astronomi	50
6	Biologi	55
7	Geografi	67
8	Ilmu Alam	33
9	Anatomi	48
10	Lainnya	10
	Jumlah	387

B. Pengolahan Dokumen

Pengolahan dokumen dilakukan dengan tahap preprocessing data agar didapatkan data yang siap diproses lebih lanjut. Tahap preprocessing meliputi tokenizing, stopword removal, dan stemming. Contoh proses dan hasil dari tahap-tahap ini berturut-turut ditunjukkan pada Tabel 2, Tabel 3, dan Tabel 4.

C. Analisis Sistem

Pada penelitian ini metode berbasis aturan (*rule based*) digunakan sebagai metode untuk mendapatkan jawaban atas pertanyaan yang dimasukkan pada sistem. *Rule based* adalah metode yang digunakan untuk memberikan *scoring* pada dokumen-dokumen relevan yang terambil dari hasil pencocokan kata kunci

Algoritma rule yang digunakan pada penelitian ini mengacu pada rule yang telah diimplementasikan oleh Anggraeny (2007) dengan melakukan penyesuaian aturan yang disesiakan dengan dokumen pengujian.

Tiga komponen utama yang bekerja untuk menemukembalikan jawaban berdasarkan kata kunci pertanyaan yang dimasukkan oleh user yaitu *question analysis*, *passage retrieval*, dan *answer extraction*.

1. Question analysis

Kata kunci pertanyaan akan diambil oleh sistem yang selanjutnya dilakukan proses tokenizing, stopword removal dan stemming. Hasil dari tahap pre-processing data pada kata kunci pertanyaan ini hanya disimpan secara temporary pada sistem selama sistem sedang menjalankan query pertanyaan tersebut untuk menemu kembalikan jawaban. Selain itu sistem akan mengambil informasi tentang jenis kata tanya yang digunakan pada query pertanyaan yaitu dapat berupa kata tanya “apa”, “siapa”, “kapan”, “di mana”. Jenis kata tanya ini akan disimpan di dalam sistem yang nantinya digunakan pada tahap ekstraksi jawaban.

2. Passage retrieval

Kata kunci yang telah dinormalisasi pada tahap question analysis selanjutnya akan dilakukan pencocokan kata terhadap data terjemah ayat yang telah disimpan pada tahap pengolahan dokumen. Hasil pencocokan kata tersebut akan menemu kembalikan dokumen-dokumen yang relevan berupa data terjemah ayat yang memiliki minimal satu kata yang sama dengan kata kunci.

Untuk mendapatkan dokumen-dokumen yang relevan tersebut proses pertama yang dilakukan adalah memberikan bobot kata pada data terjemah ayat yang

TABEL II
CONTOH HASIL TOKENIZING

Sebelum dilakukan tokenizing	Sesudah dilakukan tokenizing
Dialah Allah, yang menjadikan segala yang ada di bumi untuk kamu dan Dia (Allah) berkehendak (menciptakan) langit, lalu dijadikan Nya tujuh langit. Dan Dia (Allah) Maha Mengetahui segala sesuatu.	'dialah', 'allah', 'yang', 'menjadikan', 'segala', 'yang', 'ada', 'di', 'bumi', 'untuk', 'kamu', 'dan', 'dia', 'allah', 'berkehendak', 'menciptakan', 'langit', 'lalu', 'dijadikan', 'nya', 'tujuh', 'langit', 'dan', 'dia', 'allah', 'maha', 'mengetahui', 'segala', 'sesuatu'.

TABEL III
CONTOH HASIL PENGHAPUSAN STOPWORD

Sebelum dilakukan penghapusan stopword removal	Sesudah dilakukan penghapusan stopword
Dialah Allah, yang menjadikan segala yang ada di bumi untuk kamu dan Dia (Allah) berkehendak (menciptakan) langit, lalu dijadikan Nya tujuh langit. Dan Dia (Allah) Maha Mengetahui segala sesuatu.	'dialah', 'allah', 'menjadikan', 'bumi', 'untuk', 'kamu', 'dia', 'allah', 'berkehendak', 'menciptakan', 'langit', 'dijadikan', 'nya', 'tujuh', 'langit', 'dia', 'allah', 'mengetahui',

TABEL IV
CONTOH HASIL STEMMING

Sebelum dilakukan stemming	Sesudah dilakukan stemming
Dialah Allah, yang menjadikan segala yang ada di bumi untuk kamu dan Dia (Allah) berkehendak (menciptakan) langit, lalu dijadikan Nya tujuh langit. Dan Dia (Allah) Maha Mengetahui segala sesuatu.	'dia', 'allah', 'jadi', 'bumi', 'untuk', 'kamu', 'dia', 'allah', 'hendak', 'cipta', 'langit', 'jadi', 'nya', 'tujuh', 'langit', 'dia', 'allah', 'tahu'

pertanyaan berdasarkan aturan yang telah dibuat.

tersimpan menggunakan metode TF-IDF.

Setelah proses pembobotan selesai, kata kunci akan dicocokkan dengan hasil pembobotan TF-IDF pada data terjemah ayat di mana proses pencocokan ini nantinya menghasilkan dokumen-dokumen yang relevan berupa data terjemah ayat yang memiliki minimal satu kata yang sama dengan kata kunci.

Proses terakhir pada tahap ini yaitu sistem akan mengambil maksimal 10 data terjemah ayat teratas hasil proses *cosine similarity* pada dokumen-dokumen yang relevan yang diambil dengan kata kunci pertanyaan. Kemudian dengan menggunakan indeks dari data terjemah ayat yang diambil ini, akan dipanggil dokumen asli dan dilakukan proses *parsing* atau pemecahan paragraf menjadi kalimat-kalimat menggunakan tanda titik (.) sebagai parameternya. Dokumen asli yang telah dipisah per kalimat akan kembali dinormalisasi melalui tahap *pre-processing* data dan disimpan sementara di dalam sistem beserta indeksnya.

3. Answer extraction

Data terjemah ayat berupa kalimat-kalimat hasil proses *cosine similarity* akan dicocokkan kembali dengan kata kunci pertanyaan menggunakan algoritma *wordMatch* dan *rule based* untuk mendapatkan jawaban yang sesuai dengan tipe pertanyaannya. Algoritma *wordMatch* digunakan untuk membandingkan setiap token kalimat data terjemah ayat dengan setiap token kalimat pertanyaan. Pada saat proses perbandingan berlangsung, nilai *clue* (+3) akan ditambah jika setiap token kalimat dokumen sama dengan token pertanyaan pada kalimat data terjemah ayat yang dibandingkan.

Setelah pencocokan dengan algoritma *wordMatch* selesai selanjutnya dilakukan pemberian nilai pada kalimat data terjemah ayat menggunakan *rule based*. Suatu *rule* memiliki empat alternatif nilai yaitu *slam_dunk* (+20), *confident* (+6), *good_clue* (+4), dan *clue* (+3). Kalimat dengan nilai tertinggi akan dikembalikan sebagai jawabannya. Algoritma berbasis aturan yang digunakan pada tahap ini mengacu pada metode *rule based* yang dikembangkan oleh Anggraeny (2007) dengan melakukan beberapa modifikasi. Mengacu pada Angraeny (2007), penggunaan *rule-based* tersebut juga memerlukan pengelompokan tipe jawaban yang disusun ke dalam suatu kamus, yaitu kamus_orang, kamus_waktu, dan kamus_tempat, sedangkan untuk pertanyaan “apa” tidak menggunakan kamus.

1. Orang

“Orang” merupakan hasil jawaban dari pertanyaan dengan kata tanya “siapa”. Jawaban dari pertanyaan ini dapat diidentifikasi dengan adanya sebutan di antaranya “Allah”, “Nabi”, dan “seseorang”. Di antara kata yang termasuk dalam kamus “orang” ini adalah “Allah”, “Dia”, “Kami”, “Engkau”, “syaitan” dan “manusia”. Kamus yang digunakan

untuk menyimpan tipe jawaban untuk pertanyaan ini adalah kamus “kamus_orang”.

2. Waktu

“Waktu” merupakan hasil jawaban dari pertanyaan dengan kata tanya “kapan”. Pertanyaan ini untuk mengidentifikasi terkait waktu terjadinya suatu peristiwa. Contoh kata yang termasuk dalam jawaban dari pertanyaan ini seperti ‘hari’, ‘malam’, dan ‘tahun’. Kamus “kamus_waktu” digunakan untuk menyimpan tipe jawaban ini.

3. Tempat

“Tempat” merupakan hasil jawaban dari pertanyaan “di mana”. “Tempat” terdiri dari beberapa nama tempat yang ada pada dokumen. Contoh beberapa nama tempat dalam dokumen di antaranya ‘Shafa’, ‘neraka’, dan ‘surga’. Kamus “kamus_tempat” digunakan untuk menyimpan jawaban ini.

4. Untuk pertanyaan “apa” tidak menggunakan kamus sebagai pengelompokan tipe jawaban, namun jawaban akan langsung dicari di dalam sistem

sesuai dengan *rule* yang telah dibuat dengan cara mengidentifikasi baik itu kata yang terkandung dalam dokumen tersimpan maupun dalam *query* pertanyaan. Misalnya jika *query* pertanyaan mengandung kata “maksud” maka jawaban yang dicari pada dokumen tersimpan akan menambahkan nilai pada *rule* jika terdapat kata “adalah”, “yaitu” dan “ialah”.

Algoritma *rules* yang digunakan sebagai acuan pada sistem sebagai berikut:

Rule 1. “Siapa”

Q = (Q - {siapa, siapakah})

Score(S) += wordMatch (Q,S)

If contain(Q, orang) and contain(S, orang) then

Score(S) += slam_dunk

Rule 2. “Apa”

Q = (Q - {apa})

Score(S) += wordMatch (Q,S)

If contain(Q,{fungsi,manfaat,tujuan}) and contain (S,{agar,supaya,untuk, guna}) then

Score(S) += confident

If contain(Q,{nama,maksud,sebut,kenal}) and contain(S,{ialah,adalah,yaitu,arti}) then

Score(S) += slam_dunk

Rule 3. “Di mana”

Q = (Q - {di manakah, di mana, kemana, kemanakah})

```
Score(S) += wordMatch(Q,S)
If contain(S, {pada,dalam,dari }) and
  contain(S,TEMPAT) then
Score(S) += slam_dunk
If contain(S,TEMPAT) then
Score(S) += good_clue
If contain(S, {pada,dalam,dari }) then
Score(S) += clue
```

Rule 4. “Kapan”

```
Q = (Q - {kapan, kapankah})
Score(S) += wordMatch(Q,S)
If contain(S, {ketika,kala,saat,sejak,semenjak,
  waktu,sebelum,setelah,sesudah,selama,pada}) and
  contains(S, WAKTU) then
Score(S) += slam_dunk
If contain(S, WAKTU) and contains(Q, WAKTU)
  then
Score(S) += confident
If contain(S, {ketika,saat,kala,sejak,semenjak,
  waktu,setelah,sebelum,sesudah,selama,pada}) or
  contain(S, WAKTU) then
Score(S) += good_clue
```

Keterangan:

1. Notasi S = *sentence* (kalimat dokumen) adalah terjemahan ayat al-Qur’an yang terdapat dalam database.
2. Notasi Q = *query* (kalimat kata kunci) kalimat pertanyaan yang diberikan *user* ke sistem.
3. *contain* merupakan fungsi untuk memeriksa kalimat dokumen dan kalimat *query* pertanyaan, apakah pertanyaan tersebut mengandung kata yang telah ditentukan.
4. *wordMatch* merupakan fungsi untuk memeriksa kesamaan antar kata.
5. *score* merupakan fungsi untuk memberikan nilai pada kalimat dokumen.

V. PEMBAHASAN

A. Skenario Uji Coba

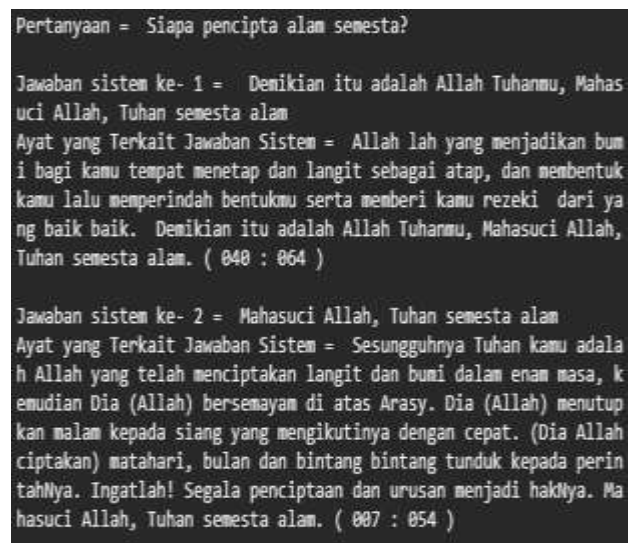
Uji coba terhadap sistem dilakukan dengan menggunakan 20 pertanyaan. Rincian dari 20 pertanyaan tersebut terdiri dari lima pertanyaan untuk masing-masing jenis kata, seperti ditunjukkan pada Tabel 5.

Untuk memverifikasi kebenaran jawaban yang dihasilkan oleh sistem, digunakan *gold standard* jawaban dari pakar sebanyak dua orang yang ahli dibidang al-Qur’an. Pakar pertama memiliki latar belakang sebagai guru iman yang mengajar di sekolah Islam swasta, sedangkan pakar kedua sebagai guru mengaji yang hafal 30 juz dan memiliki tempat

pendidikan mengaji al-Qur’an. Tahapan untuk memverifikasi jawaban sistem kepada para pakar, akan digunakan daftar pertanyaan pada Tabel 5 yang akan dijawab sesuai pemahaman pakar. Jawaban yang diberikan oleh pakar yaitu berupa ayat pada surat di al-Qur’an dengan jumlah jawaban tidak dibatasi atau boleh lebih dari satu jawaban untuk setiap pertanyaan.

B. Uji Coba Sistem

Prosedur pengujian sistem dilakukan dengan cara memasukkan satu persatu pertanyaan yang ada pada Tabel 5 ke dalam sistem. Ketika proses pencarian jawaban telah selesai sistem akan menampilkan hasil jawabannya, contoh hasil pencarian seperti pada Gambar 3. Sistem akan menampilkan pertanyaan beserta jawaban yang merujuk pada ayat dan surat keberapa didalam al-Qur’an yang ditandai dengan angka didalam tanda kurung dimana angka pertama menandakan nomor surat sedangkan dibelakang tanda (:) adalah nomor ayat. Sistem juga dapat menampilkan jawaban lebih dari satu, seperti ditunjukkan pada Gambar 3, untuk pertanyaan “Siapa pencipta alam semesta?” menghasilkan 2 jawaban, yaitu Surat 040 ayat 064 dan Surat 007 ayat 054.



Gambar 3. Hasil jawaban sistem dengan query “Siapa pencipta alam semesta?”

Nilai *precision* digunakan untuk mengukur tingkat akurasi dari proses pencarian yang dilakukan oleh sistem, nilainya didapatkan dari jumlah jawaban sistem yang sama dengan jawaban relevan pakar dibagi dengan jumlah jawaban sistem yang terambil untuk setiap pertanyaan.

Nilai *recall* digunakan untuk mengukur kemampuan pencarian dalam menemukan jawaban yang relevan, nilainya didapatkan dari jumlah jawaban sistem yang sama dengan jawaban relevan pakar dibagi dengan jumlah jawaban relevan pakar.

TABEL V
DAFTAR PERTANYAAN

#Query	Deskripsi Pertanyaan	Kata Tanya	Dokumen Relevan
1	Siapa pencipta alam semesta?	Siapa	007:054
2	Siapa yang menurunkan hujan?	Siapa	042:028; 016:010
3	Kepada siapa Allah menghembuskan angin yang sangat dingin dan membinasakan?	Siapa	054:019; 046:024
4	Siapa yang dapat memahami bahasa binatang?	Siapa	027:016; 027:018-019
5	Dari siapa Qabil belajar cara menguburkan mayat saudaranya?	Siapa	005:031
6	Manusia diciptakan dari apa?	Apa	021:020; 032:007-009; 075:037; 077:020; 076:002; 086:006-007; 023:012-014
7	Apa saja yang diharamkan dimakan oleh manusia?	Apa	006:145; 005:003
8	Untuk apa alquran diturunkan kepada manusia?	Apa	002:002; 005:016; 006:019; 016:064; 036:069; 045:020
9	Apa dasar yang digunakan untuk perhitungan waktu?	Apa	010:005
10	Laut apa yang berhasil dilintasi oleh Bani Israil dari kejaran Firaun?	Apa	002:050; 010:090
11	Di mana Musa ingin melihat Allah?	Di mana	007:143
12	Di mana anak Nabi Nuh akan mencari perlindungan dari terjan- gan banjir bandang?	Di mana	011:043
13	Di manakah lebah dapat membuat sarang sesuai petunjuk dari Tuhannya?	Di mana	016:068
14	Di mana tempat yang Allah hiasi dengan bintang-bintang?	Di mana	015:016; 037:006; 067:005
15	Saat berada di mana ratu semut khawatir terinjak oleh Nabi Sulaiman dan bala tentaranya?	Di mana	027:018
16	Kapan Nabi Muhammad melakukan perjalanan isra miraj?	Kapan	017:001
17	Kapan Allah memerintahkan nabi musa untuk memukulkan tongkatnya ke laut?	Kapan	026:063; 020:077
18	Kapan biasanya manusia menjadikan kulit binatang ternak se- sebagai rumah?	Kapan	016:080
19	Kapan Allah mengambil kesaksian dari manusia bahwa Allah adalah tuhan mereka?	Kapan	007:172
20	Pada saat kapan manusia diharamkan memburu hewan di darat?	Kapan	005:096

TABEL VI
HASIL PRECISION, RECALL, DAN F-MEASURES

#Query	Dokumen yang dihasilkan Sistem	Dokumen Relevan	Precision	Recall	F-measures
1	007:054; 040:064	007:054	0.5	1	0.67
2	015:022; 007:057; 042:028	042:028; 016:010	0.33	0.5	0.4
3	051:041-042	054:019; 046:024	0	0	0
4	027:016	027:016; 027:018-019	1	0.5	0.67
5	005:031	005:031	1	1	1
6	086:006-007	021:020; 032:007-009; 075:037; 077:020; 076:002; 086:006-007; 023:012-014	1	0.14	0.25
7	006:145; 005:003	006:145; 005:003	1	1	1
8	002:185	002:002; 005:016; 006:019; 016:064; 036:069; 045:020	0	0	0
9	010:005	010:005	1	1	1
10	007:138; 010:090	002:050; 010:090	0.5	0.5	0.5
11	007:143	007:143	1	1	1
12	011:043	011:043	1	1	1
13	016:068; 016:069	016:068	0.5	1	0.67
14	037:006; 037:007; 037:008; 037:-009; 041:012	015:016; 037:006; 067:005	0.2	0.33	0.25
15	027:018	027:018	1	1	1
16	017:001	017:001	1	1	1
17	020:077	026:063; 020:077	1	0.5	0.67
18	016:080	016:080	1	1	1
19	007:172	007:172	1	1	1
20	005:096	005:096	1	1	1
Rerata			0.75	0.72	0.74

Nilai *precision* tertinggi dengan nilai 1 terdapat pada 13 pertanyaan yang diajukan yaitu pada query 4, 5, 6, 7, 9, 11, 12, 15, 16, 17, 18, 19, 20. Kemudian untuk nilai *recall* tertinggi dengan nilai 1 terdapat pada 12 pertanyaan yang diajukan yaitu pada query 1, 5, 7, 9, 11, 12, 13, 15, 16, 18, 19, 20. Untuk nilai terendah dari *precision* dan *recall* memiliki nilai yang sama yaitu 0 yang terdapat pada 2 pertanyaan yang diajukan yaitu pada query 3 dan 8. Dari nilai *precision* dan *recall* yang didapat akan digunakan untuk menghitung nilai *f-measures*. Rata-rata nilai *precision*, *recall* dan *f-measures* untuk setiap pertanyaan yang diajukan kedalam sistem yaitu sebesar 0.75 untuk nilai *precision*, 0.72 untuk nilai *recall*, dan 0.74 untuk nilai *f-measures* seperti pada Tabel 6.

VI. KESIMPULAN

Question-answering system pada terjemahan ayat suci al-Quran berbahasa Indonesia dapat diimplementasikan menggunakan metode rule-based. Implementasi ini terdiri dari 3 tahap, yaitu *question analysis*, *passage retrieval* dan *answer extraction*. Hasil dari evaluasi sistem menggunakan 20 pertanyaan didapatkan hasil yang baik dengan rerata *precision*, *recall*, dan *f-measures* berturut-turut sebesar 0.75, 0.72, dan 0.74.

Alternatif pengembangan lebih lanjut untuk penelitian ini, di antaranya penerapan part-of-speech tagging untuk identifikasi jenis kata, penerapan thesaurus dan wordmatch untuk identifikasi kata terkait.

DAFTAR PUSTAKA

- [1] Khan, M. R. (2013). Science in the Quran. *Journal of Basic and Applied Scientific Research*, 3(1), 106–111.
- [2] Fakhry, J. (2010). Sains dan Teknologi dalam al-Qur'an dan Implikasinya dalam Pembelajaran. *Ta'dib: Jurnal Pendidikan Islam*, 15(01), 121-142.
- [3] Khusna, A. N., & Mardhia, M. M. (2019). Perancangan Aplikasi Question Answering System Pada Terjemahan Alquran. *Journal of Information System Management (JOISM)*, 1(1), 8-12.
- [4] Anggraeny, M. D. (2007). Implementasi Question Answering System Dengan Metode Rule-Based Pada Terjemahan Al Qur'an Surat Al Baqarah. Skripsi Program Studi Ilmu Komputer IPB Bogor.
- [5] Purwarianti, A., & Yusliani, N. (2011). Sistem Question Answering Bahasa Indonesia untuk Pertanyaan Non-Factoid. *Jurnal Ilmu Komputer dan Informasi*, 4(1), 10-14.
- [6] Setiarini, D., Gusmita, R. H., & Agustin, F. (2020). Sistem Tanya Jawab Menggunakan Kata Tanya Non-Factoid pada Terjemah Bahasa Indonesia Surat Al Baqarah. *Jurnal Linguistik Komputasional*, 3(1), 25-32.
- [7] Purwaningrum, S. (2015). Elaborasi ayat-ayat sains dalam Al-Quran: Langkah menuju integrasi agama dan sains dalam pendidikan. *Inovatif: Jurnal Penelitian Pendidikan, Agama, Dan Kebudayaan*, 1(1), 124-141.
- [8] Fatimah, F. (2017). Ayat-ayat Sains dalam al-Quran (Telaâah Balaghah). *Jurnal Al-Hikmah*, 5(2), 1-9.
- [9] Riloff, E., & Thelen, M. (2000). A rule-based question answering system for reading comprehension tests. In *ANLP-NAACL 2000 workshop: reading comprehension tests as evaluation for computer-based language understanding systems*.
- [10] Sudin, S., Joenadi, H., & Santosa, J. (2019). Analisis Jenis Pertanyaan Berbahasa Indonesia pada Question and Answering System Menggunakan Metode Support Vector Machine (SVM). *Dintek*, 12(1), 72-80.
- [11] Zaman, B., & Winarko, E. (2011). Analisis Fitur Kalimat untuk Peringkat Teks Otomatis pada Bahasa Indonesia. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 5(2).
- [12] Romadhony, A., Al Faraby, S., Rismala, R., Wisesty, U. N., & Arifianto, A. (2024). Sentiment Analysis on a Large Indonesian Product Review Dataset. *Journal of Information Systems Engineering and Business Intelligence*, 10(1), 167–178. <https://doi.org/10.20473/jisebi.10.1.167-178>
- [13] Zaman, B., Justitia, A., Sani, K. N., & Purwanti, E. (2020). An Indonesian hoax news detection system using reader feedback and naïve bayes algorithm. *Cybernetics and Information Technologies*, 20(1), 82-94.
- [14] Roiqoh, Salsabila, Badrus Zaman, and Kartono Kartono. "Analisis Sentimen Berbasis Aspek Ulasan Aplikasi Mobile JKN dengan Lexicon Based dan Naïve Bayes." *Jurnal Media Informatika Budidarma 7.3 (2023): 1582-1592*.
- [15] Priandini, N., Zaman, B., & Purwanti, E. (2017, August). Categorizing document by fuzzy C-Means and K-nearest neighbors approach. In *AIP Conference Proceedings (Vol. 1867, No. 1)*. AIP Publishing.
- [16] Herlambang, G. H., Nugroho, A., & Zaman, B. (2020). Klasifikasi Perkiraan Kelulusan Mahasiswa Jenjang Magister Menggunakan Metode Naive Bayes. *NJCA (Nusantara Journal of Computers and Its Applications)*, 5(1), 40-46.