

Sistem Tanya Jawab Menggunakan Kata Tanya Non-Factoid pada Terjemah Bahasa Indonesia Surat Al Baqarah

Dewi Setiari¹, Ria Hari Gusmita², Fenty Eka Muzayyana Agustin³

^{1,2,3}Program Studi teknik Informatika, Universitas Islam Syarif Hidayatullah Jakarta
Jl. Ir. H. Juanda No. 95, Ciputat, Tangerang Selatan, Banten

¹ setiariidewi7@gmail.com

²ria.gusmita@uinjkt.ac.id

³fenty.eka@uinjkt.ac.id

Abstrak— Sistem tanya jawab memiliki tiga komponen utama yaitu penganalisis pertanyaan, pengambilan dokumen relevan, dan ekstraksi jawaban. penelitian terkait dengan menggunakan kata tanya tipe non-factoid (apa, mengapa, bagaimana). Bagian penting dari sistem tanya jawab non-factoid adalah pola kalimat. Pada penelitian sistem tanya jawab non-factoid yang telah dilakukan yaitu pada dokumen bahasa Indonesia murni, sedangkan terjemahan Al Quran bahasa Indonesia adalah dokumen terjemahan. Hal tersebut memungkinkan karakteristik teks bahasa Indonesia murni dengan teks bahasa Indonesia berbeda, sehingga pada penelitian ini melakukan analisis dokumen terjemahan Al Quran bahasa Indonesia untuk menentukan pola kalimat. Sistem tanya jawab dengan menggunakan pola kalimat dari hasil analisis dokumen terjemahan bahasa Indonesia surat Al Baqarah masih belum mampu menjawab semua pertanyaan dengan benar di mana hanya 73,33% jawaban benar yang dikembalikan. Namun, kinerja sistem lebih baik dari pada sistem tanya jawab yang hanya menggunakan pola kalimat yang telah digunakan pada penelitian sebelumnya yaitu 51,11% jawaban benar yang dikembalikan.

Kata kunci— Sistem tanya jawab, Al Baqarah, non-factoid

I. PENDAHULUAN

Sistem temu kembali informasi (Information Retrieval) saat ini banyak memiliki fitur query berupa pertanyaan sehingga informasi atau jawaban yang dikembalikan kepada pengguna akan lebih cepat dalam proses penemuan masalah. Salah satu dari sistem temu kembali informasi adalah Question Answering (QA) System atau Sistem Tanya Jawab. Beberapa tipe sistem tanya jawab berdasarkan jawaban yang dihasilkan dengan tipe pertanyaan yang digunakan terbagi 5 jenis pertanyaan yaitu factoid, non-factoid, yes-no, list, dan opini [1].

Telah banyak penelitian sistem tanya jawab untuk domain bahasa Indonesia, salah satunya penelitian yang berjudul Sistem Question Answering Bahasa Indonesia

untuk Pertanyaan Non-Factoid. Penelitian tersebut merupakan penelitian sistem tanya jawab pertama untuk dokumen bahasa Indonesia yang menggunakan pertanyaan jenis non-factoid dengan motivasi bahwa pertanyaan non-factoid merupakan pertanyaan yang sering digunakan dalam usaha memahami sesuatu.

Sistem tanya jawab dapat dimanfaatkan untuk mendapatkan informasi dalam segala bidang, termasuk bidang keagamaan. Agama merupakan hal terpenting dalam kehidupan manusia. Tak terkecuali pada agama Islam. Islam mempunyai Al Quran sebagai pedoman bagi pemeluknya. Seperti yang dijelaskan dalam terjemahan Surat Al Baqarah ayat 185 yaitu: “Bulan Ramadhan adalah (bulan) yang di dalamnya diturunkan Al Quran, sebagai petunjuk bagi manusia dan penjelasan-penjelasan mengenai petunjuk itu dan pembeda (antara yang hak dan yang bathil)”.

Di Indonesia telah dilakukan beberapa penelitian tentang sistem tanya jawab pada terjemahan Al Quran. Penelitian tentang sistem tanya jawab pada terjemahan Al Quran yang sudah dilakukan hanya menggunakan kata tanya tipe factoid dan belum ada penelitian tentang sistem tanya jawab pada terjemahan Al Quran dengan memakai kata tanya tipe non-factoid [2]. Penelitian ini menerapkan pendekatan yang berbeda dalam mengidentifikasi pola kalimat yang digunakan untuk memperoleh jawaban.

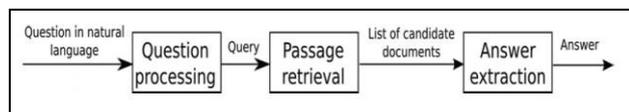
II. PENELITIAN SEJENIS

Sistem Tanya Jawab membahas masalah untuk menemukan jawaban atas pertanyaan yang diajukan dalam bahasa alami. Secara tradisional sistem tanya jawab diharapkan dapat memberikan satu jawaban singkat untuk permintaan pengguna [3]. Sistem tanya jawab bekerja dengan menggabungkan bidang IR lain seperti *indexing-searching*, *information extraction*, dan *summarization*. Sistem tanya jawab akan menintegrasikan informasi-informasi yang bersumber dari berita, artikel, buku, dan

lainnya untuk memberikan jawaban yang sesuai dengan pertanyaan dari pengguna [4].

Beberapa jenis sistem tanya jawab berdasarkan jawaban yang dihasilkan dengan tipe pertanyaan yang digunakan terbagi 5 jenis pertanyaan yaitu factoid, non-factoid, yes-no, list, dan opini. Jenis non-factoid, yaitu pertanyaan yang sering digunakan dalam usaha memahami sesuatu. Pertanyaan dibagi menjadi tiga kategori yaitu definisi, alasan, dan metode [1]

Sistem tanya jawab memiliki tiga tahapan pokok yaitu question processing, passage retrieval, dan answer extraction.



Gambar 1. Arsitektur dasar sistem tanya jawab [3]

A. Pola Kalimat Bahasa Indonesia Untuk Sistem Tanya Jawab Non-Factoid

Untuk pertanyaan non-factoid, salah satu metode yang dapat digunakan yaitu pattern matching. Dalam metode pattern matching, komponen answer extraction menggunakan aturan pola kalimat dan kata petunjuk (clue word) dari kalimat. Metode ini adalah metode monolingual sederhana untuk menemukan jawaban dari pertanyaan non-factoid. Pola kalimat yang mencakup semua kemungkinan kemunculan kata kunci dan kata petunjuk dalam kalimat. Setiap pola memiliki nilai prioritas. Jika kalimat memiliki prioritas yang lebih tinggi, semakin besar kemungkinan akan menjadi jawaban dari pertanyaan [5].

Penelitian yang dilakukan oleh Novi Yusliani dan Ayu Purwarianti yaitu membangun sistem tanya jawab non-factoid yang merupakan usaha pertama untuk membangun sistem tanya jawab non-factoid dokumen bahasa Indonesia. Pada penelitian tersebut, telah didefinisikan pola kalimat beserta urutan prioritas setiap pola kalimatnya [1]. Berikut urutan pola kalimat berdasar prioritas sebagai berikut:

- ... + kata khusus + satu atau lebih kata kunci tanpa stem
- satu atau lebih kata kunci tanpa stem + kata khusus + ...
- satu atau lebih kata kunci tanpa stem
- ... + kata khusus + satu atau lebih kata kunci dengan stem
- satu atau lebih kata kunci dengan stem + kata khusus + ...
- satu atau lebih kata kunci dengan stem

B. Korpus

Korpus adalah kumpulan dokumen yang digunakan sebagai kamus data. Pada sistem tanya jawab bahasa Indonesia, korpus juga harus dokumen dalam bahasa Indonesia dan tidak harus mencakup hanya satu topik atau

domain [5]. Pada penelitian ini, korpus yang digunakan adalah terjemahan bahasa Indonesia Surat Al Baqarah yang berasal dari Tanzil Quran. Tanzil Quran adalah proyek Quran yang diluncurkan pada awal tahun 2007. Tujuan dibuatnya Tanzil adalah untuk menghasilkan teks standar Unicode Quran dan berfungsi sebagai sumber terpercaya untuk teks standar pada web Tanzil. Tampilan teks Al Quran pada Tanzil dapat dipilih menurut surat, ayat, dan juz (Tanzil Quran).

III. ANALISIS DAN PERANCANGAN SISTEM TANYA JAWAB

A. Analisis Dokumen Terjemahan Al Baqarah

Hasil dari tahap analisis ini adalah penulis mendapatkan terjemahan ayat-ayat terpilih yang berisi tentang penjelasan suatu hal tertentu, sehingga memudahkan analisis pada tahap selanjutnya. Dengan membaca, penulis juga menemukan kesalahan-kesalahan pengetikan kata pada terjemahan Al Baqarah. Ada beberapa perbaikan dari kesalahan pengetikan yaitu :

- Pada terjemahan ayat 228, “Wanita-wanita yang ditalak hendaklah menahan diri (menunggu) tiga kali quru'.” Kata “hendaklah” diganti dengan kata hendaklah.
- Pada terjemahan ayat 282, “Dan janganlah penulis enggan menuliskannya sebagaimana Allah mengajarkannya, meka hendaklah ia menulis, dan hendaklah orang yang berhutang itu mengimlakkan (apa yang akan ditulis itu),” Kata “meka” diganti dengan kata maka.
- Pada terjemahan ayat 282, “Jika tak ada dua orang lelaki, maka (boleh) seorang lelaki dan dua orang perempuan dari saksi-saksi yang kamu ridhai,” Kata “oang” diganti dengan kata orang.

Berdasarkan penjelasan diatas, maka dokumen terjemahan Al Baqarah dari Tanzil Quran telah diperiksa kembali oleh penulis. Sehingga diharapkan tidak ada kesalahan pengetikan kata.

B. Analisis Konten Ayat dan Jenis Pertanyaan

Jenis kata tanya non-factoid yaitu apa, mengapa, dan bagaimana. Dari tiap-tiap kata tanya tersebut memiliki tipe jawaban yaitu kata tanya “apa” menjelaskan tentang definisi, kata tanya “mengapa” menjelaskan tentang alasan, dan kata tanya “bagaimana” menjelaskan tentang metode. Beberapa kata khusus yang digunakan dalam penelitian sebelumnya [1]. Penelitian ini memilih terjemahan ayat mana saja yang menjelaskan tentang definisi, alasan, atau metode. Selanjutnya ayat-ayat yang terpilih dari tahap sebelumnya, dikelompokkan menurut *Expected Answer Type (EAT)*.

TABEL I
CONTOH PENGELOMPOKAN AYAT BERDASARKAN EAT

<i>Expected Answer Type (EAT)</i>			
No.	Defini	Alasan	Metode
1.	Dan (ingatlah) ketika Kami selamatkan kamu dari (Fir'aun dan) pengikut-pengikut Fir'aun; mereka menimpakan siksaan yang sangat berat kepadamu, mereka menyembelih anak-anak laki-lakimu dan membiarkan hidup anak-anak perempuanmu. Dan pada yang demikian itu merupakan cobaan yang besar dari Tuhanmu.	Dan (ingatlah), ketika Kami berikan kepada Musa Al Kitab (Taurat) dan keterangan yang membedakan antara yang benar dan yang salah, agar kamu mendapat petunjuk.	... Maka barangsiapa diantara kamu sakit atau dalam perjalanan (lalu tidak berpuasa), maka (wajib mengganti) sebanyak hari (yang dia tidak berpuasa itu) pada hari-hari yang lain. Dan bagi orang yang berat menjalankannya, wajib membayar fidyah, yaitu: memberi makan seorang miskin. ...
2.	... apa yang ditumbuhkan bumi, yaitu sayur-mayurnya, ketimunnnya, bawang putihnya, kacang adasnya, dan bawang merahnya	Dan persaksikanlah dengan dua orang saksi dari orang-orang lelaki (di antaramu); Jika tak ada dua orang lelaki, maka (boleh) seorang lelaki dan dua orang perempuan dari saksi-saksi yang kamu ridhai, supaya jika seorang lupa maka yang seorang mengingatkannya	... Dan barangsiapa sakit atau dalam perjalanan (dia tidak berpuasa), maka (wajib menggantinya) , sebanyak hari yang ditinggalkannya itu, pada hari-hari yang lain. ...
	...sapi betina itu adalah sapi betina yang tidak tua dan tidak muda; pertengahan antara itu; maka kerjakanlah apa yang diperintahkan kepadamu.		Kepada wanita-wanita yang diceritakan (hendaklah diberikan oleh suaminya) mut'ah menurut yang ma'ruf, sebagai suatu kewajiban bagi orang-orang yang bertakwa.

Berdasarkan tabel I, dapat dilihat bahwa perbedaan dari penjelasan ayat selain pada kata khusus yang digunakan sebagai petunjuk, perbedaan juga terdapat pada isi konten ayat tersebut. Selanjutnya ditentukan kata khusus (*clue word*) dari setiap EAT. Berikut daftar kata khusus (*clue word*).

TABEL III
Kata khusus (*clue word*) untuk setiap EAT

EAT	Kata tanya	Kata khusus
Definisi	Apa Apakah	yaitu, bahwa, adalah, ialah, seperti
Alasan	Mengapa Kenapa	agar, supaya, karena
Metode	Bagaimana Bagaimanakah	barangsiapa, jika, apabila, maka, hendaklah

C. Menentukan Pola Kalimat

Pola kalimat mengacu pada penelitian sebelumnya [1], [5]. Namun dalam terjemahan Al Quran bahasa Indonesia mempunyai karakter kalimat yang berbeda dengan karakter teks bahasa Indonesia yang ditulis oleh manusia, sehingga penulis mendefinisikan beberapa pola kalimat yang berbeda dari penelitian sebelumnya. Dari semua tahap analisis, maka penelitian ini mendefinisikan 8 pola kalimat, yakni (1) semua kata kunci tanpa *stem* + kata khusus, (2) satu atau lebih kata kunci tanpa *stem* + kata khusus, (3) khusus + semua kata kunci tanpa *stem*, (4) kata khusus + satu atau lebih kata kunci tanpa *stem*, (5) semua kata kunci tanpa *stem*, (6) satu atau lebih kata kunci tanpa *stem*, (7) kata khusus+semua kata kunci tanpa *stem*+kata khusus, (8) kata khusus+satu atau lebih kata kunci tanpa *stem*+kata khusus.

D. Arsitektur Sistem Tanya Jawab

Arsitektur sistem tanya jawab memiliki tiga komponen utama yaitu (1) penganalisis pertanyaan, (2) pengambil dokumen relevan, dan (3) ekstraksi jawaban. Tahapan pertama yaitu user memberikan masukan berupa pertanyaan berbahasa Indonesia ke dalam sistem. Pertanyaan diproses oleh penganalisis pertanyaan yang menghasilkan kata kunci dan tipe jawaban atau *Expected Answer Type (EAT)*. Tahapan kedua yaitu pengambilan dokumen relevan dengan mencari dokumen relevan pada korpus yang telah tersedia. Dokumen yang diambil adalah dokumen yang mengandung kata kunci sehingga akan menjadi kandidat jawaban. Dan tahap terakhir adalah ekstraksi jawaban yang akan menghasilkan jawaban dari pertanyaan yang diberikan oleh user.

Pertanyaan yang diberikan oleh user akan masuk ke dalam modul Penganalisis Pertanyaan. Modul akan menganalisa pertanyaan untuk menentukan tipe jawaban (EAT) dan mendapatkan kata kunci. Tipe jawaban akan diperoleh dengan cara melakukan string matching pada kalimat pertanyaan. Dan jika ditemukan kata tanya yang sesuai maka modul akan menyimpan tipe jawaban (EAT) yang sesuai dengan kata tanya tersebut. Pertanyaan yang

masuk akan diperiksa oleh modul dengan melakukan pencocokkan kata dengan menggunakan metode string matching. Jika kalimat pertanyaan mengandung kata “apa” atau “apakah” maka tipe jawaban (EAT) adalah definisi, jika mengandung kata “mengapa” atau “kenapa” maka tipe jawaban (EAT) adalah alasan, dan jika mengandung kata “bagaimana” atau “bagaimanakah” maka tipe jawaban (EAT) adalah metode. Contoh pertanyaan yang diberikan oleh user:

“Apa saja yang tumbuh di bumi?”

Dari pertanyaan tersebut modul akan mendapatkan kata tanya “Apa”, dimana kata tanya “Apa” menunjukkan EAT “definisi”. Modul akan menyimpan definisi sebagai EAT dari pertanyaan tersebut. Tahap selanjutnya, modul akan mencari kata kunci dari pertanyaan. Kata kunci dapat berupa kata benda, kata kerja, atau kata sifat. Modul akan melakukan tahap-tahap sebagai berikut; (1) menghapus kata tanya “apa”, (2) menghapus tanda tanya (?), (3) menghapus stopword “saja” dan “yang”. Pada tahap ekstraksi query akan menghasilkan kata, “tumbuh bumi” yang akan menjadi query.

Tahap pengambilan dokumen relevan bertujuan untuk mencari dokumen relevan pada korpus yang telah tersedia. Dokumen yang diambil adalah dokumen yang mengandung kata kunci sehingga akan menjadi kandidat jawaban. Kata kunci diperoleh dari tahapan ekstraksi query, yaitu query. Pencarian kandidat jawaban akan dilakukan dengan string matching antara query dan korpus yang tersedia. Dokumen yang terambil yaitu dokumen yang mengandung semua atau salah satu kata kunci. Dokumen tersebut akan menjadi kandidat jawaban untuk proses ekstraksi jawaban.

Pada tahap terakhir dari proses sistem tanya jawab yaitu ekstraksi jawaban, akan menghasilkan jawaban dari pertanyaan yang diberikan kepada sistem. Setelah tahap pengambilan dokumen relevan menghasilkan dokumen relevan yang merupakan kandidat jawaban, kalimat-kalimat tersebut akan diproses untuk mendapatkan jawaban yang tepat. Proses tersebut dilakukan dengan cara mencocokkan kalimat dengan pola kalimat yang telah ditentukan oleh prioritas pola yang paling tinggi, maka jawaban yang memenuhi pola dengan prioritas tertinggi akan menjadi jawaban dari sistem. Pada sebelumnya tidak ada penjelasan untuk mendefinisikan aturan prioritas atau nilai prioritas pada pola kalimat. Oleh sebab itu, pada penelitian ini penulis mendefinisikan nilai prioritas berdasarkan jumlah kemunculan setiap pola kalimat. Selain nilai prioritas, penulis juga mendefinisikan skor sebagai nilai yang mewakili setiap pola kalimat. Skor akan digunakan dalam proses ekstraksi jawaban. Berikut urutan pola kalimat berdasarkan prioritas dan skor yang ditentukan.

TABEL III
PRIORITAS DAN SKOR POLA KALIMAT

Pri orit as	Pola Kalimat	Jumlah Kemun culan	Skor
1	semua kata kunci tanpa stem+kata khusus	17	0,33
2	kata khusus+semua kata kunci tanpa stem+kata khusus	16	0,31
3	kata khusus+satu atau lebih kata kunci tanpa stem+kata khusus	5	0,10
4	satu atau lebih kata kunci tanpa stem	4	0,08
5	satu atau lebih kata kunci tanpa stem+kata khusus dan semua kata kunci tanpa stem	3	0,06
6	kata khusus+semua kata kunci tanpa stem	2	0,04
7	kata khusus+satu atau lebih kata kunci tanpa stem	1	0,02
Jumlah kemunculan semua pola kalimat		51	1

Berdasarkan tabel di atas nilai prioritas adalah kode pola kalimat yang memiliki skor tertinggi. Semakin tinggi skor semakin menjadi prioritas yang lebih utama. Dari dokumen yang relevan dipecah menjadi kalimat, sehingga pencocokan pola kalimat dilakukan pada setiap dokumen. Dalam pencocokan kalimat akan mengambil data EAT dan kata khusus (*clue word*) untuk setiap EAT. Kalimat yang menjadi jawaban adalah kalimat yang memenuhi pola dengan prioritas tertinggi.

E. Pengujian Sistem

Pengujian sistem dilakukan dengan mencoba 45 pertanyaan yang dibuat oleh penulis dimana semua pertanyaan memiliki jawaban pada dokumen terjemahan Al Baqarah. Pertanyaan tersebut dibagi tiga jenis yaitu 22 pertanyaan dengan tipe jawaban definisi, 5 pertanyaan dengan tipe jawaban alasan, dan 18 pertanyaan dengan tipe jawaban metode. Hasil pengujian sistem pada penelitian ini, penulis mendapatkan kategori jawaban, yaitu benar, tidak tepat, dan salah. Pada terjemahan Al Baqarah tidak semua ayat hanya terdiri dari satu kalimat. Namun banyak ditemukan bahwa dalam satu ayat terdiri dari dua atau lebih kalimat yang saling berkaitan. Setiap terjemahan ayat terpilih diberikan pertanyaan yang sesuai dengan jenis jawaban.

Dari pertanyaan tersebut dapat diperoleh kata kunci yang kemudian dicocokkan dengan dokumen. Jika dokumen memiliki semua kata kunci atau satu atau lebih kata kunci yang ada maka dokumen tersebut memenuhi salah satu pola kalimat, dimana syarat dari pola kalimat

yaitu memiliki kata kunci. Kemudian dokumen akan diperiksa kembali apakah memiliki kata khusus yang telah ditentukan sebelumnya atau tidak.

TABEL IVV
DAFTAR PERTANYAAN

No.	Pertanyaan
1.	Allah mengharamkan makanan apa?
2.	Apa balasan orang yang kembali mengambil riba?
3.	Apa berita gembira bagi mereka yang beriman dan berbuat baik?
4.	Apakah haidh itu?
5.	Apa hukum berperang di bulan Haram?
6.	Mengapa Allah timpakan musibah dari langit kepada orang zalim?
7.	Mengapa diperlukan saksi dua orang saksi dalam berhutang?
8.	Mengapa kita diwajibkan berpuasa?
9.	Mengapa Kitab Taurat diberikan kepada Musa?
10.	Mengapa orang yang makan riba seperti gila?
11.	Bagaimana jika bermu'amalah tidak tunai dalam perjalanan?
12.	Bagaimana jika bermu'amalah tidak tunai untuk waktu yang ditentukan?
13.	Bagaimana jika dalam keadaan takut atau bahaya?
14.	Bagaimana jika dalam perjalanan saat berpuasa lalu berbuka ?
15.	"Seperti apa orang yang beriman kepada Al Kitab?"

Kategori jawaban dari sistem tanya jawab terdiri enam macam [6], yang ada di tabel berikut.

TABEL V
KATEGORI JAWABAN

Kategori	Keterangan
Benar	Sistem mengembalikan jawaban yang benar dan diperoleh dari dokumen yang relevan
Tidak Mendukung	Sistem mengembalikan jawaban benar namun diperoleh dari dokumen yang tidak relevan dengan pertanyaan
Tidak Tepat	Sistem mengembalikan jawaban berupa string yang lebih dari jawaban yang sebenarnya, ataupun kehilangan sebagian dari jawaban dan diperoleh dari dokumen yang relevan
Salah	Sistem mengembalikan jawaban yang salah
Tidak terjawab	Sistem tidak dapat menemukan jawaban

IV. HASIL DAN PEMBAHASAN

Hasil pengujian sistem pada penelitian ini, penulis mendapatkan kategori jawaban, yaitu benar, tidak tepat, dan salah. Berikut adalah hasil pengujian sistem tanya jawab.

TABEL V
HASIL PENGUIAN SISTEM BERDASARKAN KATEGORI

Kategori	Jumlah	Prosentase (%)
Benar	25	55,56
Tidak Mendukung	0	0,00
Tidak Tepat	11	24,44
Salah	9	20,00
Tidak terjawab	0	0,00
	45	100,00

Berikut ini adalah tabel contoh pertanyaan (*query*) dan jawaban yang diberikan oleh sistem disertai skornya.

TABEL VII
CONTOH PERTANYAAN DAN JAWABAN

No.	Pertanyaan	Jawaban	Skor	Kategori
1	Apakah haidh itu?	mereka bertanya kepadamu tentang haidh. katakanlah: "haidh itu adalah suatu kotoran". oleh sebab itu hendaklah kamu menjauhkan diri dari wanita di waktu haidh; dan janganlah kamu mendekati mereka, sebelum mereka suci. apabila mereka telah suci, maka campurilah mereka itu di tempat yang diperintahkan allah kepadamu. Sesungguhnya allah menyukai orang-orang yang bertaubat dan menyukai orang-orang yang mensucikan diri. (Ayat 222)	0,06	Benar
2	Apa balasan orang yang kembali mengambil riba?	maka jika kamu tidak mengerjakan (meninggalkan sisa riba), maka ketahuilah, bahwa allah dan rasulnya akan memerangimu. dan jika kamu bertaubat (dari pengambilan riba), maka bagimu pokok hartamu; kamu tidak menganiaya dan tidak (pula) dianiaya. (Ayat 279)	0,08	Salah

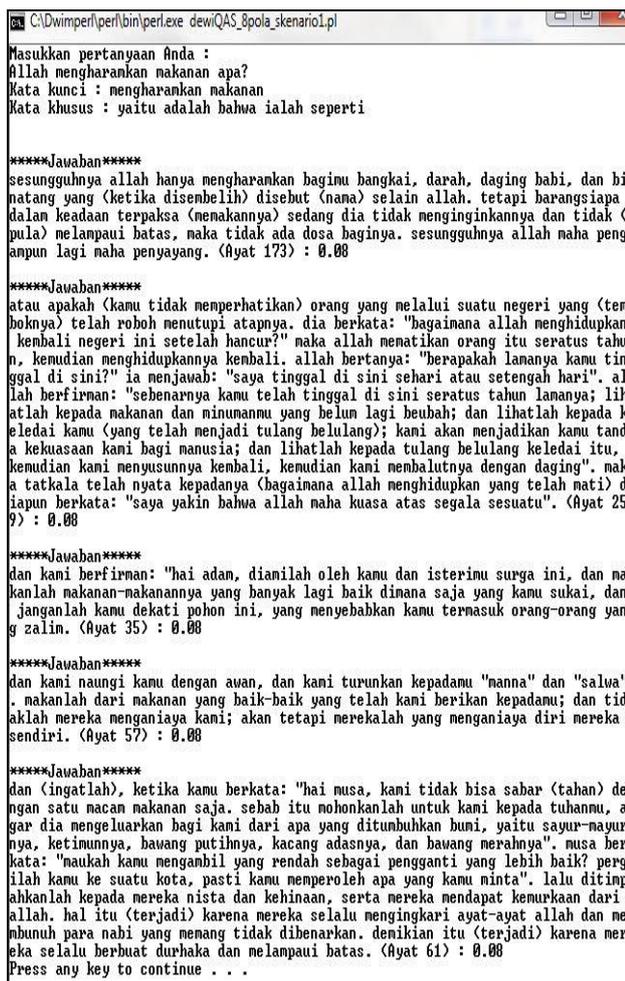
Jawaban tidak tepat yaitu jika sistem mengembalikan jawaban berupa *string* yang lebih dari jawaban yang sebenarnya, ataupun kehilangan sebagian dari jawaban dan diperoleh dari dokumen yang relevan. Dari 45

pertanyaan, jumlah pertanyaan yang dijawab tidak tepat adalah 11 pertanyaan.

Pada pertanyaan “Allah mengharamkan makanan apa?” sistem mengembalikan 5 jawaban, yaitu dokumen terjemahan ayat 173, 259, 35, 57, dan 61.

Jawaban benar adalah dokumen terjemahan ayat 173, yang berisi:

sesungguhnya allah hanya mengharamkan bagimu bangkai, darah, daging babi, dan binatang yang (ketika disembelih) disebut (nama) selain allah. tetapi barangsiapa dalam keadaan terpaksa (memakannya) sedang dia tidak menginginkannya dan tidak (pula) melampaui batas, maka tidak ada dosa baginya. sesungguhnya allah maha pengampun lagi maha penyayang. (Ayat 173)

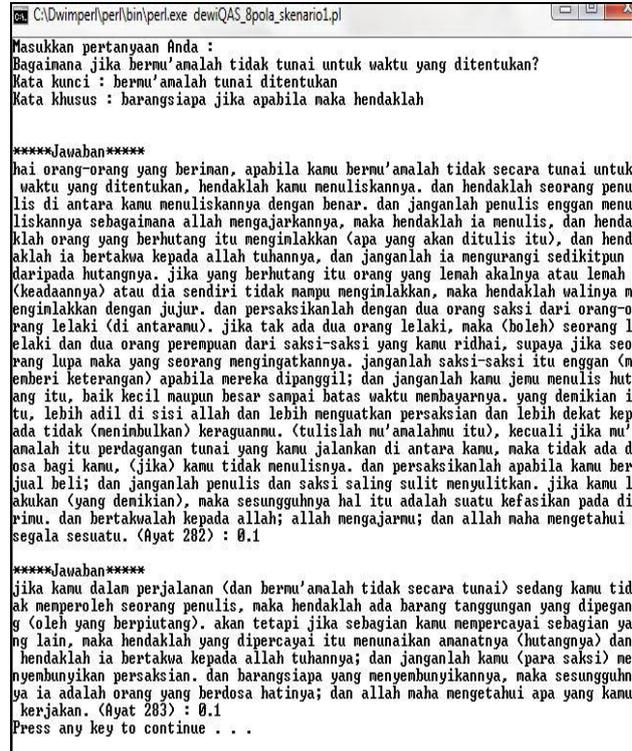


Gambar 1. Pertanyaan ke-1, Jawaban dan Skornya

Dokumen tersebut memenuhi pola kalimat “satu atau lebih kata kunci tanpa stem” dengan skor pola kalimat 0,08 yang merupakan skor tertinggi dari semua kandidat jawaban. 4 ayat yang lain (259, 35, 57 dan 61) juga memiliki skor yang sama (0.08), hal ini mengakibatkan

sistem mengembalikan dokumen jawaban lain selain dokumen jawaban ayat 173.

Pada pertanyaan “Bagaimana jika bermu’amalah tidak tunai untuk waktu yang ditentukan?” sistem mengembalikan 2 jawaban, yaitu dokumen terjemahan ayat 282 dan 283 yang masing-masing memiliki skor 0,1.



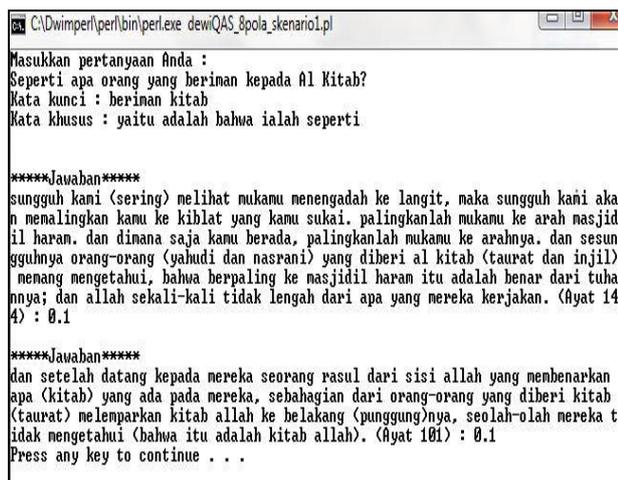
Gambar 2. Pertanyaan ke-11, Jawaban dan Skornya

Dokumen tersebut memenuhi pola kalimat “kata khusus+satu atau lebih kata kunci tanpa stem+kata khusus” dengan skor pola kalimat 0,1 yang merupakan skor tertinggi dari semua kandidat jawaban.

Sedangkan ayat 283 meskipun skornya tinggi (0,1), tapi pola kalimatnya tidak sesuai dengan yang sudah ditetapkan. Oleh karena itu, jawaban dari sistem ini termasuk kategori jawaban tidak tepat.

Selanjutnya dianalisa jenis pertanyaan dengan kategori jawaban salah. Pada pertanyaan “Seperti apa orang yang beriman kepada Al Kitab?” sistem mengembalikan jawaban dokumen terjemahan ayat 101 dan 144 dengan skor 0,1 yang memenuhi pola kalimat “kata khusus+satu atau lebih kata kunci tanpa stem+kata khusus”. Dalam hal ini jawaban benar adalah dokumen terjemahan ayat 121, yang berisi:

orang-orang yang telah kami berikan al kitab kepadanya, mereka membacanya dengan bacaan yang sebenarnya, mereka itu beriman kepadanya. dan barangsiapa yang ingkar kepadanya, maka mereka itulah orang-orang yang rugi. (Ayat 121)



Gambar 3. Pertanyaan ke-15, Jawaban dan Skornya

Dokumen yang merupakan jawaban benar memenuhi pola kalimat “semua kata kunci tanpa stem” dengan skor yang lebih rendah dari skor dokumen jawaban sistem yaitu 0,08. Sedangkan dokumen terjemahan ayat 101 dan 144 merupakan salah satu kandidat dokumen yang memiliki skor tertinggi dengan skor 0,08 sehingga sistem mengembalikan dokumen tersebut sebagai jawaban. Dengan demikian, jawaban ini dimasukkan sebagai jawaban dengan kategori salah.

Selanjutnya untuk mengetahui kinerja sistem pada penelitian ini, maka dilakukan evaluasi pada sistem dengan melakukan pengujian yang menggunakan pola kalimat yang ada pada penelitian sebelumnya, tanpa menggunakan pola kalimat baru yang ditemukan pada penelitian ini dari hasil analisis terjemahan bahasa Indonesia Surat Al Baqarah dan menggunakan skor yang sama telah ditentukan pada penelitian ini. Berikut pola kalimat yang digunakan [1], [5]: (1) semua kata kunci tanpa stem+kata khusus, (2) satu atau lebih kata kunci tanpa stem+kata khusus, (3) kata khusus+semua kata kunci tanpa stem, (4) kata khusus+satu atau lebih kata kunci tanpa stem, (5) semua kata kunci tanpa stem, (6) satu atau lebih kata kunci tanpa stem.

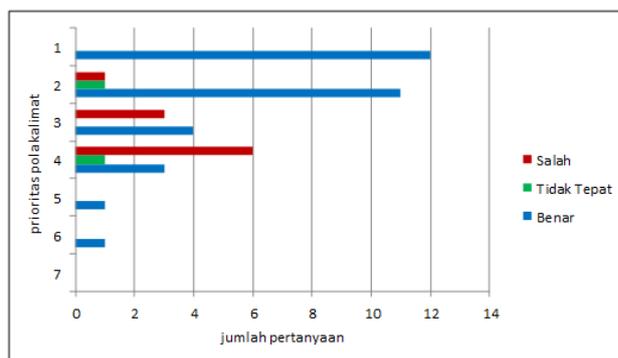
Berikut adalah hasil dari pengujian sistem tanpa menggunakan pola kalimat baru

TABEL V
HASIL PENGUJIAN SISTEM BERDASARKAN KATEGORI

Kategori	Jumlah	Prosentase (%)
Benar	23	51,11
Tidak Mendukung	0	0,00
Tidak Tepat	1	2,22
Salah	21	46,67
Tidak terjawab	0	0,00
	45	100,00

Berdasarkan tabel di atas, hasil dari pengujian sistem tanpa menggunakan pola kalimat baru yaitu sistem menghasilkan 51,11% jawaban benar, 2,22% jawaban tidak tepat, dan 46,67% jawaban salah. Sedangkan pada penelitian ini dimana sistem menggunakan pola kalimat berdasarkan hasil analisis dokumen terjemahan bahasa Indonesia Surat Al Baqarah menghasilkan 73,33% jawaban benar, 4,44% jawaban tidak tepat, dan 22,22% jawaban salah. Dengan demikian dapat disimpulkan bahwa penentuan pola kalimat berdasarkan hasil analisis dokumen terjemahan bahasa Indonesia Surat Al Baqarah dapat meningkatkan kinerja sistem.

Selain mengetahui kinerja dari sistem, penelitian ini juga melakukan pengecekan pada setiap penggunaan pola kalimat yang ditemukan pada analisis terjemahan bahasa Indonesia Surat Al Baqarah. Berikut ini grafik dari penggunaan pola kalimat dengan prioritas tertinggi yang dikembalikan sebagai jawaban.



Gambar 4. Grafik skor setiap pola kalimat

Berdasarkan grafik di atas, pola kalimat baru yaitu pola kalimat prioritas 2 dan 3, tidak semuanya memberikan kontribusi terhadap sistem untuk memberikan jawaban benar. Terbukti bahwa pola kalimat yang memberikan jawaban benar terbanyak adalah pola kalimat prioritas 1 dan 2, dimana pola kalimat prioritas 2 merupakan pola kalimat yang baru dari hasil yang ditemukan pada analisis terjemahan bahasa Indonesia Surat Al Baqarah. Sedangkan pola kalimat yang lebih banyak memberikan jawaban salah yaitu pada pola kalimat prioritas 3 dan 4 dimana pola kalimat prioritas 3 merupakan pola kalimat yang baru dari hasil yang ditemukan pada analisis terjemahan bahasa Indonesia Surat Al Baqarah. Namun pola kalimat prioritas 4 yaitu “satu atau lebih kata kunci tanpa stem” lebih banyak memberikan kontribusi jawaban salah.

V. KESIMPULAN

Membangun sistem tanya jawab dimulai dengan menganalisis dokumen terjemahan bahasa Indonesia surat Al Baqarah. Analisis dilakukan untuk menemukan pola kalimat dan kata khusus yang terdapat pada terjemahan surat Al Baqarah. Komponen penganalisis pertanyaan bertujuan untuk mengambil kata kunci dari setiap pertanyaan dan menentukan tipe jawaban atau EAT. Komponen pengambilan dokumen relevan bertujuan untuk mencari dokumen relevan berdasarkan kata kunci yang telah diperoleh dari penganalisis pertanyaan. Dan untuk komponen ekstraksi jawaban bertujuan untuk mendapatkan jawaban dari dokumen-dokumen terpilih dengan mencocokkan pola-pola kalimat yang telah didefinisikan sebelumnya.

Sistem tanya jawab dengan menggunakan pola kalimat dari hasil analisis dokumen terjemahan bahasa Indonesia surat Al Baqarah masih belum mampu menjawab semua pertanyaan dengan benar di mana hanya 73,33% jawaban benar yang dikembalikan. Namun, kinerja sistem lebih baik dari pada sistem tanya jawab yang mendefinisikan pola kalimat tanpa menganalisis dokumen yang akan digunakan. Kinerja sistem yang hanya menggunakan pola kalimat yang telah digunakan pada penelitian sebelumnya yaitu 51,11% jawaban benar yang dikembalikan.

Pola kalimat yang lebih banyak memberikan jawaban benar terbanyak adalah pola kalimat prioritas 1 dan 2 dimana pola kalimat prioritas 2 merupakan pola kalimat yang baru dari hasil yang ditemukan pada analisis terjemahan bahasa Indonesia Surat Al Baqarah. Sedangkan pola kalimat yang lebih banyak memberikan jawaban salah yaitu pada pola kalimat prioritas 3 dan 4 dimana pola kalimat prioritas 3 merupakan pola kalimat yang baru dari hasil yang ditemukan pada analisis terjemahan bahasa Indonesia Surat Al Baqarah.

Berdasarkan hasil analisis pada penelitian ini, diketahui bahwa pola kalimat prioritas 4 yaitu “satu atau lebih kata

kunci tanpa stem” merupakan penyebab sistem mengembalikan jawaban salah terbanyak. Hal ini dikarenakan pola tersebut memungkinkan dokumen yang dianggap sebagai jawaban hanya dengan memiliki satu kata kunci. Di sisi lain, dokumen hanya dengan memiliki satu kata kunci faktanya tidak dapat mewakili semua kata kunci dari pertanyaan pengguna.

REFERENSI

- [1] A. Purwarianti and N. Yusliani, “Sistem Question Answering Bahasa Indonesia untuk Pertanyaan Non-Factoid,” *J. Ilmu Komput. dan Inf.*, vol. 4, no. 1, pp. 10–14, 2012.
- [2] Y. Kurniawati and P. P. A. Indriati, “Implementasi named entity recognition pada factoid question answering system untuk cerita rakyat indonesia,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput. e-ISSN*, vol. 2548, p. 964X, 2018.
- [3] C. Jouis, *Next Generation Search Engines: Advanced Models for Information Retrieval: Advanced Models for Information Retrieval*. IGI Global, 2012.
- [4] S. Büttcher, C. L. A. Clarke, and G. V Cormack, *Information retrieval: Implementing and evaluating search engines*. Mit Press, 2016.
- [5] A. A. Zulen and A. Purwarianti, “Study and implementation of monolingual approach on indonesian question answering for factoid and non-factoid question,” in *Proceedings of the 25th Pacific Asia Conference on Language, Information and Computation*, 2011, pp. 622–631.
- [6] Z. Naf’an and R. H. Gusmita, “Development of an Indonesian Question Answering System about Khulafaur Rasyidin’s History,” in *Proceedings of The 1st International Conference on Cyber & IT Service Management in Conjunction with the ITIL v. 3 Workshop, Training and Certification*.