

Penerapan Merger Retriever pada Question Answering System Hadits

Rozi Saputra¹, Nazruddin Safaat Harahap², Novriyanto³, Muhammad Affandes⁴

¹Universitas Islam Negeri Sultan Syarif Kasim Riau, 12050117172@students.uin-suska.ac.id, Jl. HR. Soebrantas No.Km. 15, RW.15, Simpang Baru, Kota Pekanbaru, Riau, Indonesia

² Universitas Islam Negeri Sultan Syarif Kasim Riau, nazruddin.safaat@uin-suska.ac.id, Jl. HR. Soebrantas No.Km. 15, RW.15, Simpang Baru, Kota Pekanbaru, Riau, Indonesia

³Universitas Islam Negeri Sultan Syarif Kasim Riau, Novriyanto@uin-suska.ac.id, Jl. HR. Soebrantas No.Km. 15, RW.15, Simpang Baru, Kota Pekanbaru, Riau, Indonesia

⁴Universitas Islam Negeri Sultan Syarif Kasim Riau, affandes@uin-suska.ac.id, Jl. HR. Soebrantas No.Km. 15, RW.15, Simpang Baru, Kota Pekanbaru, Riau, Indonesia

Informasi Makalah

Submit : Apr 20, 20xx
Revisi : May 20, 20xx
Diterima : May 30, 20xx

Kata Kunci :

Question Answering System
LangChain
Gemini-pro
Merger Retriever
Large Language Model

Abstrak

Dalam dunia digital saat ini, teknologi informasi telah memberikan dampak signifikan terhadap berbagai aspek kehidupan, termasuk dalam studi dan penelusuran hadits. Penelitian ini bertujuan untuk mengembangkan *question answering system* hadits yang memungkinkan pengguna mendapatkan jawaban yang cepat dan akurat dari pertanyaan-pertanyaan terkait hadits. Dengan menggunakan *LangChain* dan model *Gemini-pro*, serta menerapkan metode *merger retriever*, sistem ini memberikan solusi inovatif dalam pencarian hadits dari 9 kitab utama (*kutubut tis'ah*). *LangChain* dan *Gemini-pro* memastikan bahwa jawaban yang dihasilkan tidak hanya relevan tetapi juga akurat, dengan memanfaatkan teknologi *Large Language Model* (LLM). Selain itu, penerapan *merger retriever* meningkatkan kinerja sistem dalam menemukan jawaban yang paling sesuai dari berbagai sumber hadits secara bersamaan. Penelitian ini melibatkan serangkaian tahapan mulai dari pengumpulan data hadits, analisa dan perancangan sistem, implementasi, hingga pengujian sistem. Hasil pengujian menunjukkan bahwa *question answering system* Hadits dapat menghasilkan jawaban yang relevan dan akurat, dengan evaluasi kualitas jawaban oleh responden ahli hadits menunjukkan rata-rata kepuasan sebesar 88,5% dengan interval penilaian "Sangat Setuju" dan *scoring evaluator LangChain* dengan rata-rata 91% serta interval termasuk dalam kategori "Sangat Setuju". Hal ini menekankan potensi besar penerapan teknologi informasi dalam studi hadits, serta membuka peluang bagi pengembangan lebih lanjut dalam bidang ini. Penelitian ini diharapkan memberikan kontribusi penting untuk kemudahan akses informasi hadits yang akurat dan cepat bagi umat Islam.

Abstract

In today's digital world, information technology has significantly impacted various aspects of life, including the study and exploration of hadith. This research aims to develop a hadith question answering system that enables users to obtain quick and accurate answers to questions related to hadith. By utilizing LangChain and the Gemini-pro model, and implementing the merger retriever method, this system provides an innovative solution in searching hadith from the 9 main books (*kutubut tis'ah*). LangChain functions as a bridge between the Large Language Model (LLM) and the hadith question answering system. Gemini-pro is a generative AI language model used to process and generate answers to user questions. Additionally, the implementation of the merger retriever enhances the system's performance in finding the most appropriate answers from various hadith sources simultaneously. This research goes through stages of hadith data collection, system analysis and design, implementation, and system testing. The test results show that the Hadith QA System can produce relevant and accurate answers, with the quality evaluation of answers by 4 hadith expert respondents showing an average satisfaction rate of 88.5% and a LangChain evaluator scoring that assesses the system's answer accuracy against hadith reference sources with an average result of 91%. Based on the Likert scale, testing with hadith experts falls under the "Strongly Agree" category, and the LangChain evaluator scoring indicates that, on average, 91% of the answers to 10 questions match the reference sources used.

1. Pendahuluan

Hadis adalah sumber penting dalam ajaran Islam yang merupakan kumpulan informasi yang berasal dari Nabi Muhammad SAW., termasuk kata-kata, tindakan, persetujuan, serta atribut fisik dan mentalnya. Hadis dianggap sebagai sumber kedua setelah Al-Qur'an dalam ajaran Islam, mencakup aspek-aspek sebelum dan sesudah kenabiannya (Andariati, 2020). Al-Qur'an tidak dapat dipisahkan dari hadis karena ayat-ayatnya bersifat global dan umum. Hadis berperan sebagai penjelasan untuk Alquran, dan karenanya, keduanya tidak dapat dipisahkan sejak zaman Nabi SAW (Ummah, 2019). Kitab hadis yang paling terkenal terdiri dari 9 kitab yang disusun oleh 9 Imam berbeda yang meliputi Shahih Muslim yang disusun oleh Muslim, Shahih Bukhari yang disusun oleh Bukhari, Sunan at-Tirmidzi yang disusun oleh At-Tirmidzi, Sunan Abu Dawud yang disusun oleh Abu Dawud, Sunan Ibnu Majah yang disusun oleh Ibnu Majah, Muwatta Malik yang disusun oleh Imam Malik, Sunan an-Nasa'i yang disusun oleh an-Nasa'i, Sunan

Darimi yang disusun oleh Ad-Darimi, dan Musnad Ahmad yang disusun oleh Imam Ahmad bin Hambal (Ni'mah & Arifin, 2020).

Kitab-kitab hadis tersebut disebut dengan *kutubut tis'ah* yang merupakan kitab hadis yang sering dijadikan sebagai rujukan baik untuk keperluan dakwah maupun sebagai bahan pembelajaran dan jumlah *kutubut tis'ah* lebih dari 62.000 hadis, sehingga memerlukan biaya yang tidak sedikit untuk memiliki kitab-kitab hadis tersebut dan waktu yang tidak sedikit untuk mencari dan menganalisa sebuah hadis (Hafidz et al., 2022). Saat ini, kemajuan teknologi informasi dan komunikasi telah menginspirasi umat Islam untuk mengubah kitab suci dan literatur Islam menjadi bentuk digital. Al-Qur'an kini telah tersedia dalam format digital lengkap dengan terjemahan, rekaman suara, dan penafsiran dari ulama terkemuka. Namun, digitalisasi hadis agak tertinggal dibandingkan dengan Al-Qur'an, terutama mengingat karakteristik unik hadis dan jumlahnya yang lebih banyak daripada Al-Qur'an (Ummah, 2019).

Perkembangan teknologi informasi saat ini mendorong munculnya berbagai produk baru berupa hadis dalam format digital, seperti PDF atau aplikasi khusus berupa ensiklopedi hadis seperti Maktabah Syamilah atau Maktabah Alfiah, dan aplikasi 'Satu Hari Satu Hadis' berbasis android. Aplikasi ini menyajikan hadis dengan dua cara yaitu secara otomatis dengan notifikasi harian berdasarkan tema dan sub-tema, serta secara manual melalui kolom pencarian (Hadi, 2020).

Belum ada aplikasi hadis yang mengimplementasikan *question answering system* (Soares & Parreiras, 2020) hingga saat ini. Sejumlah penelitian telah dilakukan dalam domain *question answering system* dan pengolahan bahasa alami (Alanazi et al., 2021), namun belum secara spesifik mempertimbangkan penerapan LangChain (Devaraj et al., 2023) dan *Gemini-pro* dalam konteks hadis. Penelitian-penelitian sebelumnya cenderung lebih terfokus pada domain lain seperti kesehatan (Mutabazi et al., 2021), keuangan (Intelligence and Neuroscience, 2023), atau teks yang umumnya lebih luas daripada hadis. Kondisi ini membuka peluang untuk pengetahuan lebih lanjut terkait penerapan teknologi terkini dalam konteks hadis yang lebih spesifik.

Penelitian oleh (Ioannidis et al., 2023) membahas mengenai pemanfaatan AI generatif untuk menangani masalah kepatuhan hukum dan regulasi. AI tersebut menggabungkan kerangka kerja LangChain, GPT-4 dari OpenAI, embedding teks, dan teknik rekayasa prompt untuk mengurangi halusinasi dan menghasilkan output domain yang andal dan akurat.

Selain itu, penelitian lainnya oleh (Finardi et al., 2024) membahas penerapan *Retrieval Augmented Generation (RAG)* yang difokuskan untuk bahasa Portugis Brasil. Penelitian tersebut menggunakan model *Gemini-pro* dan *LangChain* serta membandingkan kinerjanya dengan model

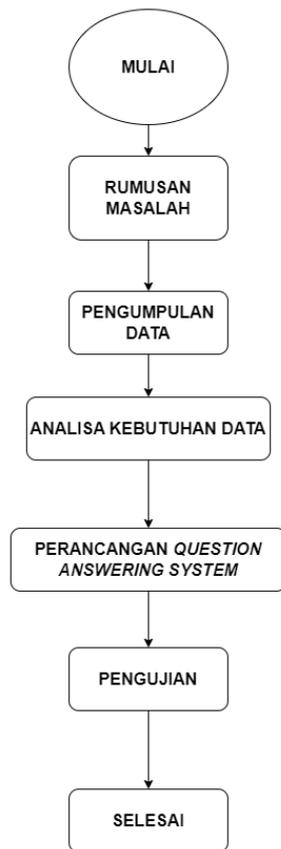
lain seperti GPT-4 dalam kemampuannya untuk menggabungkan informasi yang diambil dan menghasilkan respons yang koheren dan kontekstual.

Pada penelitian ini penulis akan merancang sebuah *question answering system* berbasis web untuk kitab-kitab hadis yang menggunakan LangChain dan Model *Gemini-pro* dengan menerapkan metode *merger retriever* (LangChain, 2023). Sistem ini dimaksudkan untuk memudahkan pengguna dalam menemukan serta memahami informasi dari hadis-hadis dengan lebih efektif dan tepat. *Question Answering System* merupakan sistem yang bertujuan untuk memberikan jawaban yang tepat dari pertanyaan pengguna, sistem ini membutuhkan kemampuan untuk menemukan jawaban dari jumlah data yang besar, baik yang terstruktur maupun yang tidak terstruktur (Alami et al., 2023).

Untuk mengoptimalkan *question answering system* hadis dengan menggunakan *merger retriever* yang menggabungkan hasil *retriever* dari 9 kitab hadis yang ada. *Retriever* adalah komponen sistem yang menghasilkan pertanyaan pencarian berbahasa alami untuk mengambil dokumen yang relevan dari mesin pencarian teks, membantu menjawab pertanyaan dengan akurasi lebih tinggi. (Qi et al., 2021).

2. Metode Penelitian

Tahapan penelitian yang akan dilakukan pada penelitian ini adalah sebagai berikut:



Gambar 1 Tahapan Penelitian

2.1. Rumusan Masalah

Masalah yang diidentifikasi dalam penelitian ini adalah bagaimana mengembangkan dan menerapkan sistem question answering (QA) berbasis LangChain dan Gemini-pro yang mampu mengakses dan memproses informasi dari sembilan kitab hadis utama (kutubut tis'ah) secara efektif. Penelitian ini bertujuan untuk:

1. Menyediakan akses cepat dan akurat terhadap informasi hadis berdasarkan pertanyaan pengguna, mengatasi keterbatasan dari metode pencarian tradisional yang sering memerlukan waktu dan usaha yang signifikan.
2. Mengintegrasikan teknologi LangChain dan Gemini-pro untuk mengubah teks hadis menjadi vektor yang dapat diproses, memungkinkan pencarian semantik yang lebih efisien dan relevan.
3. Mengoptimalkan penggunaan merger retriever untuk menggabungkan hasil pencarian dari berbagai kitab hadis, memastikan jawaban yang paling relevan dan komprehensif.

4. Mengevaluasi performa sistem melalui uji coba dengan pengguna dan ahli hadis, dengan fokus pada kepuasan pengguna, akurasi, dan kecepatan respon.

2.2. Pengumpulan Data

Data hadis pada penelitian ini diambil dari GitHub dengan pertimbangan bahwa sumber ini menyediakan koleksi hadis yang lengkap dan dapat diakses secara bebas, tepatnya dari

<https://github.com/irsyadulibad/hadis-database>. Data hadis yang digunakan sebanyak 62.169 hadis dari 9 kitab hadis yaitu kitab shahih muslim berjumlah 5.362 hadis, kitab shahih bukhari berjumlah 7.008 hadis, 3.891 hadis pada kitab sunan at-tirmidzi, kitab sunan abu dawud berjumlah 4.590 hadis, 4.332 hadis dari kitab sunan ibnu majah, 1.594 hadis dari kitab muwatta malik, kitab sunan an-nasa'i berjumlah 5.662 hadis, 3.367 hadis dari kitab sunan darimi, dan kitab musnad ahmad berjumlah 26.363 hadis. Pemilihan kitab-kitab ini didasarkan pada statusnya sebagai kutubut tis'ah. Data hadis ini disimpan dalam Database MySQL. Setelah itu data diekspor menjadi format txt agar dapat dilakukan proses selanjutnya yaitu embedding. Setiap hadis mencakup informasi berikut yaitu nomor hadis, judul hadis, teks asli dalam bahasa Arab, serta terjemahan bahasa Indonesia.

2.3. Analisa Kebutuhan Data

Data yang dibutuhkan berupa data hadis dalam bentuk vektor karena pencarian dengan menggunakan *semantic search* (Kasenchak, 2019) hanya bisa membaca data berupa data vektor. Sehingga diperlukan proses *embedding* untuk mengubah data menjadi vektor. Proses *embedding* dilakukan dengan menggunakan *LangChain* dan model *text-embedding-3-large*.

2.4. Perancangan *Question Answering*

Perancangan sistem dimulai setelah data diubah menjadi bentuk vektor dan disimpan di *ChromaDB* yang merupakan *database* yang dapat menyimpan data berbasis vektor. *Question answering system* ini menerapkan *Large Language Model* (LLM) sebagai model bahasa (Leippold, 2023). Model bahasa yang digunakan adalah *Gemini-pro* yang dirilis oleh *Google DeepMind* (Akter et al., 2023) melalui *Vertex AI* (Choi et al., 2023) (Google Cloud, 2023). Dalam makalah asli *Google* tentang *Gemini*, dilaporkan bahwa *Gemini-pro* mencapai kinerja yang mirip dengan *GPT 3.5 Turbo* (Team, 2024). Agar dapat menggunakan LLM, *question answering system* hadis memanfaatkan *LangChain*. *LangChain* merupakan *framework* yang berfungsi sebagai penghubung antara LLM dan *question answering system* hadis (Tejaswini et al., 2023). *Gemini-pro* dan *LangChain* diimplementasikan secara bersamaan untuk mengolah pertanyaan atau jawaban karena *gemini-pro* yang merupakan model bahasa dapat diimplementasikan pada *question answering system* melalui *LangChain* sebagai penghubungnya.

Sistem ini berupa web yang disimpan dan dijalankan melalui *hugging face*. *Hugging face* merupakan *platform open-source* yang berguna untuk membuat dan menyimpan suatu program kodingan seperti halnya *GitHub*. Namun perbedaannya, pada *hugging face* sistem dapat dijalankan secara langsung di *platform* tersebut. Implementasinya berupa sistem diunggah ke *hugging face*, yang menyediakan infrastruktur untuk menjalankan model NLP secara langsung di *platform* tersebut.

2.5. Pengujian

Pengujian dilakukan dengan menggunakan evaluasi kualitas jawaban sistem melalui kuesioner kepada 4 orang ahli hadis dan *scoring evaluator LangChain*. Kuesioner oleh 4 orang ahli hadis berisi 10 pertanyaan serta jawaban dari sistem yang mencakup beberapa aspek penilaian yaitu apakah jawaban yang diberikan sesuai dengan pertanyaan yang diajukan, apakah jawaban

sesuai dengan hadis yang menjadi referensi, serta apakah jawaban disajikan dengan cara yang mudah dimengerti oleh pengguna. Setiap ahli hadis memberikan skor penilaian pada skala 1-5. Sehingga nantinya para ahli hadis menilai jawaban dari sistem apakah sudah memenuhi aspek diatas sesuai dengan skala penilaian.

Pengujian lainnya yaitu dengan menggunakan *scoring evaluator LangChain*. Pengujian ini menilai jawaban sistem dengan menggunakan model bahasa dalam skala 1-10 (LangChain, 2024). Pengujian ini akan membandingkan jawaban, sehingga nantinya model bahasa akan menilai kesesuaian jawaban dengan referensi berdasarkan kriteria tertentu. Kriteria tersebut sudah disediakan di *website LangChain* pada halaman *scoring evaluator LangChain*. Adapun kriteria penilaiannya adalah sebagai berikut:

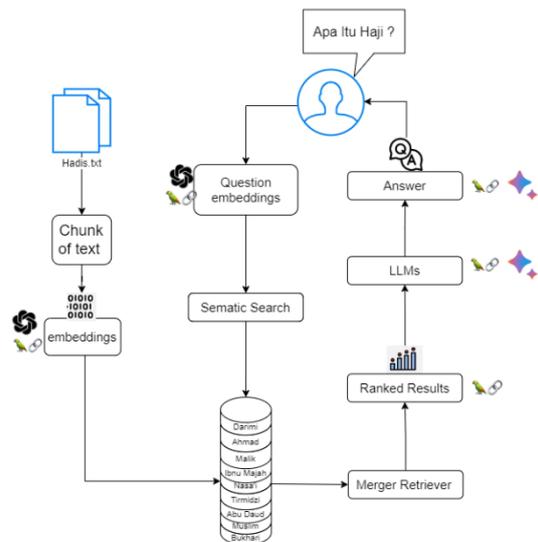
- A. Skor 1: Jika jawaban sama sekali tidak terkait dengan referensi.
- B. Skor 3: Jika jawaban memiliki relevansi minor tetapi tidak sejalan dengan referensi.
- C. Skor 5: Jika jawaban memiliki relevansi sedang tetapi mengandung ketidakakuratan.
- D. Skor 7: Jika jawaban sejalan dengan referensi tetapi memiliki kesalahan atau kelalaian minor.
- E. Skor 10: Jika jawaban benar-benar akurat dan sejalan dengan referensi dengan sempurna.

3. Hasil dan Pembahasan

Berdasarkan rancangan tersebut, maka penelitian ini masuk ke tahap selanjutnya yaitu implementasi sistem dan pengujian *question answering system* Hadis.

3.1 Implementasi Sistem

Penelitian ini menghasilkan sebuah *question answering system* Hadis. Sistem dari penelitian memanfaatkan *ChromaDB* sebagai *database* (Şorecãu & Şorecãu, 2023), *Large Language Model* (LLM) sebagai model bahasa, *LangChain* sebagai *framework* untuk menggunakan LLM (Pesaru et al., 2023), *Gemini-pro* sebagai model generatif AI.



Gambar 2. Alur Sistem

Implementasi dari sistem ini dimulai dari pengumpulan data hadis. Data hadis tersebut disimpan dengan format txt. Kemudian melakukan pembagian data mejadi beberapa *chunk* atau bagian-bagian. Hal ini dilakukan karena jumlah hadis dari setiap kitab sangatlah banyak sehingga perlu dilakukan pembagian ini guna mengoptimalisasi pada saat melakukan *embedding* nantinya. Selain itu, *chunk* juga dilakukan karena adanya keterbatasan *memory* dari model yang penulis gunakan. Pembagian *chunk* dilakukan per id hadis. Pembagian *chunk* dilakukan dengan menggunakan *library* re.

Setelah pembagian *chunk*, selanjutnya proses *embedding* data Hadis. *Embedding* adalah mengubah data hadis menjadi data vektor(Sabbeh & Fasihuddin, 2023). Hal ini dilakukan karena pencarian dengan menggunakan *semantic search*(Kasenchak, 2019) hanya bisa membaca data berupa data vektor. Proses *embedding* dilakukan dengan menggunakan *LangChain* dan model *text-embedding-3-large*. Pada *embedding* ini terjadi proses *encode* dan *decode*. Proses *encode* mengubah data teks menjadi data vector dengan menggunakan model *text-embedding-3-large*. Setiap kata dalam teks dijadikan sebagai representasi vektor dalam ruang dimensi yang besar, yang merefleksikan makna semantik kata tersebut.

Setelah *encode*, proses berikutnya adalah *decode* yang mengubah data vektor hasil *encode* menjadi teks kembali. Tujuannya agar mempertahankan teks asli meskipun sudah menjadi data vektor. Jika proses *embedding* telah selesai maka data 9 kitab hadis tersebut akan tersimpan ke *ChromaDB* dengan data berbasis vektor. Berikut ini merupakan tampilan sistem.

Question Answering Hadis

Assalamu'alaikum

Selamat datang di Sistem Question Answering Hadis! Apa yang Anda ingin tanyakan?

saya sedang berpuasa namun tidak sengaja meminum air, bagaimana dengan puasa saya?

Thinking...

Your question

Gambar 3. Tampilan Sistem

Saat memulai, sistem menampilkan judul "*Question Answering Hadis*" dan memberikan salam. Kemudian, pengguna dapat memasukkan pertanyaannya dalam kolom yang disediakan dan mengirimnya dengan meng-klik *enter*. Lalu sistem akan memproses pertanyaan tersebut dan dimulai dengan melakukan *embedding* pertanyaan menjadi bentuk vektor menggunakan *LangChain*(Topsakal & Akinci, 2023) dan model *text-embedding-3-large* untuk melakukan *embedding*. Sistem kemudian melakukan pencarian dengan *semantic search* ke data hadis yang tersimpan di *ChromaDB*(Hillebrand et al., 2023)(Arefeen et al., 2023)(Nigam et al., 2023). *Merger retriever* nantinya akan menggabungkan hasil *retriever-retriever* dari 9 kitab hadis yang menerapkan pencarian data dengan tipe similarity atau kemiripan data dengan pertanyaan pengguna. Dengan menggunakan *merger retriever* memungkinkan sistem untuk melakukan *perankingan* hadis yang dapat dijadikan referensi untuk jawaban pengguna diambil dari tiap kitab hadis. Contohnya pada sistem ini jumlah hadis yang diambil sebagai referensi untuk jawaban pengguna adalah 4 hadis dengan *rank*

tertinggi yang memiliki kemiripan atau kesesuaian dari pertanyaan pengguna. Jika menggunakan *merger retriever*, maka hadis yang akan dijadikan referensi diambil dari setiap kitab hadis. Sehingga tiap kitab hadis akan mengambil 4 hadis dengan *rank* tertinggi dari 9 kitab hadis. Dengan demikian, jumlah hadis yang dijadikan referensi untuk jawaban pengguna pada sistem ini maksimal 36 hadis. Sedangkan jika hanya menggunakan *retriever* maka hadis yang dapat dijadikan referensi dari 9 kitab hadis yang digunakan pada sistem ini hanyalah 4 hadis dengan *rank* tertinggi. Sehingga hanya beberapa kitab hadis yang data hadisnya dijadikan referensi untuk jawaban dari pertanyaan pengguna.

Lalu sistem mengurutkan jawaban yang didapatkan dari *ChromaDB* dari *rank* tertinggi. Setelah itu, sistem akan membuat atau mengolah jawaban dari pertanyaan pengguna berdasarkan hasil perankingan menggunakan *Large Language Model (LLM)* dengan model Bahasa yang digunakan adalah *Gemini-pro*. Setelah membuat jawaban, selanjutnya sistem akan mengeluarkan atau menampilkan jawaban tersebut ke pengguna.



Gambar 4 Tampilan Sistem Ketika Menampilkan Jawaban

Pada tampilan diatas terlihat bahwa ketika sistem telah selesai mencari jawaban pengguna, sistem menampilkan jawaban dan diikuti dengan tombol “Clear Chat History”. Tombol *Clear Chat Hostory* berfungsi untuk menghapus riwayat atau *history* pertanyaan pengguna selama sesi berlangsung.

3.2 Pengujian Sistem

Selanjutnya pengujian sistem dengan melakukan evaluasi kualitas jawaban dan *Scoring Evaluator LangChain*. Evaluasi kualitas jawaban dilakukan guna mengetahui apakah jawaban dari *question answering system* hadis ini sudah benar atau tidak. Pengujian evaluasi kualitas jawaban dilakukan kepada 4 orang ahli hadis. Dalam kuesioner evaluasi kualitas jawaban, terdapat beberapa opsi jawaban dan bobot penilaian yang disediakan sesuai yang tertera dalam tabel di bawah. Bobot penilaian tersebut akan digunakan sebagai faktor pengali terhadap jumlah jawaban untuk menghitung nilai total.

Tabel 1 Bobot Nilai

Tabel Bobot Nilai Kuesioner		
Pilihan Jawaban	Keterangan	Bobot Nilai
SS	Sangat Sesuai	5
S	Sesuai	4
N	Netral/Cukup	3
TS	Tidak Sesuai	2
STS	Sangat Tidak Sesuai	1

Berikut ini merupakan daftar pertanyaan kuesioner bagi pengguna beserta jawaban pertanyaan tersebut:

Tabel 2 Pertanyaan dan Jawaban Kuesioner

Tabel Isi Pertanyaan dan Jawaban Kuesioner		
No	Pertanyaan	Jawaban
1	Jika saya menemukan barang di jalan, apakah saya boleh mengambilnya?	Jika menemukan barang di jalan, dianjurkan untuk mengumumkannya selama satu tahun. Jika ada yang merasa memiliki, maka wajib dikembalikan. Namun jika tidak ada yang mengakuinya, maka barang tersebut menjadi milik penemu.
2	Ketika saya sedang berpuasa, saya tidak sengaja	Ketika berpuasa, jika tidak sengaja makan atau

Tabel Isi Pertanyaan dan Jawaban Kuesioner		
No	Pertanyaan	Jawaban
	meminum air karena saya lupa, bagaimana dengan puasa saya? apakah puasa saya akan diterima?	minum karena lupa, maka tetaplah lanjutkan puasa. Allah telah memberimu makan dan minum. Puasa tetap diterima.
3	Saya ketiduran pada pagi hari sehingga ketinggalan shalat subuh, bagaimana dengan shalat subuh saya?	Jika Anda tertidur saat waktu Subuh dan ketinggalan mengerjakan shalat, kaffarahnya adalah segera melaksanakan shalat Subuh ketika teringat. Karena tidak ada tebusan lain selain melaksanakan shalat tersebut ketika ingat.
4	Saat sedang berpuasa saya muntah, bagaimana dengan puasa saya?	Jika Anda muntah tanpa disengaja saat berpuasa, Anda tidak wajib mengganti puasa Anda. Jika Anda muntah dengan sengaja, maka Anda wajib mengganti puasa Anda.
5	Pada saat shalat saya ingin membuang angin, namun karena sedang shalat saya menahannya, bagaimana dengan shalat saya?	Jika ragu setelah mengeluarkan angin atau tidak, maka tidak perlu berwudhu hingga yakin benar. Jika ada suara atau tercium bau, maka wajib berwudhu dan mengulang shalat. Jika tidak ada, maka tidak perlu berwudhu dan melanjutkan shalat.
6	Saya memberi makan orang yang berpuasa, seperti apa pahala yang saya dapatkan?	Jika kita memberi makan orang yang sedang berpuasa, maka kita akan mendapatkan pahala yang serupa dengan pahala yang didapat oleh orang yang berpuasa tersebut. Dan pahala ini tidak akan mengurangi pahala orang yang berpuasa sedikit pun.
7	Ketika saya sedang shalat, kemudian ada orang yang mengucapkan salam kepada saya, apa yang harus saya lakukan?	Dalam hadis, disebutkan bahwa ketika sedang shalat, Rasulullah SAW tidak menjawab salam karena shalat membutuhkan kekhusyukan dan konsentrasi yang tinggi. Namun, ada hadis lain yang menjelaskan bahwa Rasulullah SAW membalas salam dengan isyarat tangan.
8	Saya sedang melaksanakan sholat dan kemudian ada orang yang lewat didepan saya, apa yang harus saya lakukan?	Jika seseorang melewatimu saat salat dan menolak untuk menghindar, maka lawanlah dia karena dia adalah setan. Hal ini sesuai sabda Rasulullah SAW yang mengatakan bahwa orang yang lewat di depan orang yang sedang salat adalah setan.
9	Ketika saya sedang duduk, tiba-tiba ada anjing yang mendekat dan menjilati kaki saya sehingga air liur anjing tersebut mengenai kaki saya, bagaimana saya membersihkannya?	Apabila seekor anjing menjilati bejana milik salah seorang dari kalian, maka hendaklah ia menumpahkan isinya, kemudian mencucinya tujuh kali, dan salah satunya dengan tanah.
10	Apa keutamaan bulan Ramadhan?	Bulan Ramadhan adalah bulan yang penuh berkah dimana pintu-pintu surga dibuka, pintu-pintu neraka ditutup, dan setan-setan dibelenggu. Pada bulan ini terdapat satu malam yang lebih baik dari seribu bulan, yaitu lailatul qadar. Selain itu, bulan Ramadhan juga merupakan bulan dimana umat Islam diwajibkan untuk berpuasa.

Berikut ini merupakan hasil evaluasi kualitas jawaban dari 4 orang ahli hadis dimana

penghitungan berdasarkan bobot nilai pada tabel 1 diatas:

Tabel 3 Hasil Evaluasi Kualitas Jawaban

Tabel Hasil Evaluasi Kualitas Jawaban								
No.	Pertanyaan	Jumlah Jawaban					Total Nilai	Persentase Nilai
		SS	S	N	TS	STS		
1	Jika saya menemukan barang di jalan, apakah saya boleh mengambilnya?	2	2	0	0	0	90	90%

Tabel Hasil Evaluasi Kualitas Jawaban

No.	Pertanyaan	Jumlah Jawaban					Total	Persentase
2	Ketika saya sedang berpuasa, saya tidak sengaja meminum air karena saya lupa, bagaimana dengan puasa saya? apakah puasa saya akan diterima?	2	2	0	0	0	90	90%
3	Saya ketiduran pada pagi hari sehingga ketinggalan shalat subuh, bagaimana dengan shalat subuh saya?	2	2	0	0	0	90	90%
4	Saat sedang berpuasa saya muntah, bagaimana dengan puasa saya?	2	2	0	0	0	90	90%
5	Pada saat shalat saya ingin membuang angin, namun karena sedang shalat saya menahannya, bagaimana dengan shalat saya?	2	2	0	0	0	90	90%
6	Saya memberi makan orang yang berpuasa, seperti apa pahala yang saya dapatkan?	2	2	0	0	0	90	90%
7	Ketika saya sedang shalat, kemudian ada orang yang mengucapkan salam kepada saya, apa yang harus saya lakukan?	2	2	0	0	0	90	90%
8	Saya sedang melaksanakan sholat dan kemudian ada orang yang lewat didepan saya, apa yang harus saya lakukan?	2	1	0	1	0	80	80%
9	Ketika saya sedang duduk, tiba-tiba ada anjing yang mendekat dan menjilati kaki saya sehingga air liur anjing tersebut mengenai kaki saya, bagaimana saya membersihkannya?	2	1	1	0	0	85	85%
10	Apa keutamaan bulan Ramadhan?	2	2	0	0	0	90	90%
Total						885	88,5%	

Berdasarkan evaluasi kualitas jawaban *question answering system* hadis kepada ahli hadis dapat diketahui bahwa nilai terendah pada soal nomor 8 dengan persentase 80% dan nilai tertinggi pada soal nomor 1 sampai 7 serta soal nomor 10 dengan persentase 90%, sedangkan rata-rata nilai *question answering system* hadis pada penelitian ini adalah 88,5%. Berikut ini merupakan interval penilaian pada bobot skala *likert*:

- Indeks 0% – 19,99% : Sangat Tidak Setuju
- Indeks 20% – 39,99% : Tidak Setuju
- Indeks 40% – 59,99% : Kurang Setuju
- Indeks 60% – 79,99% : Setuju
- Indeks 80% – 100% : Sangat Setuju

Berdasarkan skala *likert* diatas, dapat diketahui bahwa hasil evaluasi kualitas jawaban *question answering system* hadis kepada ahli hadis dengan nilai indeks 88,5% masuk kedalam kategori “Sangat Setuju”.

Pengujian lainnya yaitu dengan menggunakan *scoring evaluator LangChain*. Pengujian ini menilai jawaban sistem dengan

menggunakan model bahasa dalam skala 1-10 (LangChain, 2024). Pengujian ini akan membandingkan jawaban, sehingga nantinya model bahasa akan menilai kesesuaian jawaban dengan referensi berdasarkan kriteria tertentu. Adapun kriteria penilaiannya adalah sebagai berikut:

- A. Skor 1: Jika jawaban sama sekali tidak terkait dengan referensi.
- B. Skor 3: Jika jawaban memiliki relevansi minor tetapi tidak sejalan dengan referensi.
- C. Skor 5: Jika jawaban memiliki relevansi sedang tetapi mengandung ketidakakuratan.
- D. Skor 7: Jika jawaban sejalan dengan referensi tetapi memiliki kesalahan atau kelalaian minor.
- E. Skor 10: Jika jawaban benar-benar akurat dan sejalan dengan referensi dengan sempurna.

Berikut adalah hasil pengujian dengan *scoring evaluator LangChain*:

Tabel 4 Hasil Pengujian Scoring Evaluator LangChain

Tabel Hasil Pengujian Scoring Evaluator LangChain			
No	Pertanyaan	Skor	Persentase
1	Jika saya menemukan barang di jalan, apakah saya boleh mengambilnya?	7	70%
2	Ketika saya sedang berpuasa, saya tidak sengaja meminum air karena saya lupa, bagaimana dengan puasa saya? apakah puasa saya akan diterima?	10	100%
3	Saya ketiduran pada pagi hari sehingga ketinggalan shalat subuh, bagaimana dengan shalat subuh saya?	10	100%
4	Saat sedang berpuasa saya muntah, bagaimana dengan puasa saya?	10	100%
5	Pada saat shalat saya ingin membuang angin, namun karena sedang shalat saya menahannya, bagaimana dengan shalat saya?	7	70%
6	Saya memberi makan orang yang berpuasa, seperti apa pahala yang saya dapatkan?	10	100%
7	Ketika saya sedang shalat, kemudian ada orang yang mengucapkan salam kepada saya, apa yang harus saya lakukan?	10	100%
8	Saya sedang melaksanakan sholat dan kemudian ada orang yang lewat didepan saya, apa yang harus saya lakukan?	10	100%
9	Ketika saya sedang duduk, tiba-tiba ada anjing yang mendekat dan menjilati kaki saya sehingga air liur anjing tersebut mengenai kaki saya, bagaimana saya membersihkannya?	7	70%
10	Apa keutamaan bulan Ramadhan?	10	100%
Total		91	91%

Berdasarkan tabel diatas, dapat diketahui bahwa hasil rata-rata pengujian dengan *scoring evaluator LangChain* adalah 91%. Adapun interval penilaian pada bobot skala *likert* yaitu:

- Indeks 0% – 19,99% : Sangat Tidak Setuju
- Indeks 20% – 39,99% : Tidak Setuju
- Indeks 40% – 59,99% : Kurang Setuju
- Indeks 60% – 79,99% : Setuju
- Indeks 80% – 100% : Sangat Setuju

Nilai indeks dari *scoring evaluator LanChain* adalah 91%. Berdasarkan skala *likert* diatas, dapat bahwa interval hasil *scoring evaluator LanChain* adalah “Sangat Setuju”.

4. Simpulan

Penelitian ini mengembangkan *question answering system* berbasis hadis yang memanfaatkan *merger retriever, LangChain*, dan model *Gemini-pro*. Sistem ini dirancang untuk memudahkan pengguna dalam menemukan informasi hadis secara akurat dan cepat menggunakan *database ChromaDB* dan *Large Language Model (LLM)*. Evaluasi dilakukan dengan dua metode yaitu kuesioner kepada 4 orang ahli hadis yang menghasilkan nilai rata-rata

keseluruhan 88,5% yang menunjukkan bahwa responden “Sangat Setuju” terhadap kualitas jawaban dari sistem dan pengujian menggunakan *scoring evaluator LangChain* yang mencatat nilai rata-rata 91% dengan interval “Sangat Setuju”, sehingga menunjukkan tingkat kesesuaian yang tinggi antara jawaban sistem dan referensi. Hasil penelitian menunjukkan *question answering system* hadis berhasil menggabungkan data dari 9 kitab hadis dengan efektif.

5. Referensi

- Alami, H., El Mahdaouy, A., Benlahbib, A., En-Nahnahi, N., Berrada, I., & Ouatik, S. E. A. (2023). DAQAS: Deep Arabic Question Answering System based on duplicate question detection and machine reading comprehension. *Journal of King Saud University - Computer and Information Sciences*, 35(8). <https://doi.org/10.1016/j.jksuci.2023.101709>
- Andariati, L. (2020). HADIS DAN SEJARAH PERKEMBANGANNYA.

- Diroyah: Jurnal Ilmu Hadis*, 4(2), 153–166.
- Arefeen, M. A., Debnath, B., & Chakradhar, S. (2023). LeanContext: Cost-Efficient Domain-Specific Question Answering Using LLMs. *ArXiv*, 1–8. <http://arxiv.org/abs/2309.00841>
- Finardi, P., Avila, L., Castaldoni, R., Gengo, P., Larcher, C., Piau, M., Costa, P., & Caridá, V. (2024). *The Chronicles of RAG: The Retriever, the Chunk and the Generator*. <http://arxiv.org/abs/2401.07883>
- Hadi, R. T. (2020). STUDI APLIKASI HADIS ERA MOBILE (STUDI “SATU HARI SATU HADIS” OLEH PUSAT KAJIAN HADIS). *ISLAM TRANSFORMATIF: Journal of Islamic Studies*, 4(1), 13–24.
- Hafidz, P., Jurusan, M., Hadis, I., Ushuluddin, F., Sunan, U., & Djati Bandung, G. (2022). ANALISIS APLIKASI ENSIKLOPEDIA HADIS 9 IMAM VERSI ANDROID SEBAGAI SARANA DAKWAH DI MEDIA SOSIAL. *Jurnal Penelitian Ilmu Ushuluddin*, 2(2), 312–330. <https://doi.org/10.15575/jpiu.v2i2.14393>
- Hillebrand, L., Berger, A., Deußer, T., Dilmaghani, T., Khaled, M., Kliem, B., Loitz, R., Pielka, M., Leonhard, D., Bauckhage, C., & Sifa, R. (2023, August 22). Improving Zero-Shot Text Matching for Financial Auditing with Large Language Models. *DocEng 2023 - Proceedings of the 2023 ACM Symposium on Document Engineering*. <https://doi.org/10.1145/3573128.3609344>
- Ioannidis, J., Harper, J., Quah, M. S., & Hunter, D. (2023, June 19). Gracnote.ai: Legal Generative AI for Regulatory Compliance. *Proceedings of the Third International Workshop on Artificial Intelligence and Intelligent Assistance for Legal Professionals in the Digital Workplace*.
- Kasenchak, R. T. (2019). What is Semantic Search? and why is it important? *Information Services and Use*, 39(3), 205–213. <https://doi.org/10.3233/ISU-190045>
- LangChain. (2023). *LOTR (Merger Retriever)* |  Langchain. LangChain. https://python.langchain.com/docs/integrations/retrievers/merger_retriever
- LangChain. (2024). *Scoring Evaluator*. LangChain.
- Nigam, S. K., Mishra, S. K., Mishra, A. K., Shallum, N., & Bhattacharya, A. (2023). Legal Question-Answering in the Indian Context: Efficacy, Challenges, and Potential of Modern AI Models. *ArXiv*, 1–15. <http://arxiv.org/abs/2309.14735>
- Ni'mah, A. T., & Arifin, A. Z. (2020). Perbandingan Metode Term Weighting terhadap Hasil Klasifikasi Teks pada Dataset Terjemahan Kitab Hadis. *Rekayasa*, 13(2), 172–180. <https://doi.org/10.21107/rekayasa.v13i2.6412>
- Pesaru, A., Gill, T. S., & Tangella, A. R. (2023). AI assistant for document management Using Lang Chain and Pinecone. *International Research Journal of Modernization in Engineering Technology and Science*, 5(6), 3980–3983. <https://doi.org/10.56726/irjmets42630>
- Qi, P., Lee, H., Manning, C. D., & Sido, O. “TG.” (2021). Answering Open-Domain Questions of Varying Reasoning Steps from Text. *Proceedings Of the 2021 Conference on Empirical Methods in Natural Language Processing*, 3599–3614. <https://beerqa.github.io/>
- Sabbeh, S. F., & Fasihuddin, H. A. (2023). A Comparative Analysis of Word Embedding and Deep Learning for Arabic Sentiment Classification. *Electronics (Switzerland)*, 12(6). <https://doi.org/10.3390/electronics12061425>
- Șorecău, M., & Șorecău, E. (2023). An Alternative Application to CHATGPT that Uses Reliable Sources to Enhance the Learning Process. *International Conference KNOWLEDGE-BASED ORGANIZATION*, 29(3), 113–119. <https://doi.org/10.2478/kbo-2023-0084>
- Topsakal, O., & Akinci, T. C. (2023). Creating Large Language Model Applications Utilizing LangChain: A Primer on Developing LLM Apps Fast. *5th International Conference on Applied*

Engineering and Natural Sciences,
1050–1056. [http://as-
proceeding.com/:Konya,Turkeyhttps://w
ww.icaens.com/](http://as-proceeding.com/:Konya,Turkeyhttps://www.icaens.com/)

Ummah, S. S. (2019). DIGITALISASI
HADIS (Studi Hadis di Era Digital).
Diroyah: Jurnal Ilmu Hadis, 4(1), 1–10.