



Application of Langchain Technology to the Fiqh Question Answering System of Four Madhhab

Penerapan Teknologi LangChain pada Question Answering System Fikih Empat Madzhab

Suci Rahayu^{1*}, Nazruddin Safaat Harahap², Surya Agustian³, Pizaini⁴

^{1,2,3,4}Program Studi Teknik Informatika, Fakultas Sains dan Teknologi,
Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia

E-Mail: ¹sucirahayu2097@gmail.com, ²nazruddin.safaat@uin-suska.ac.id,
³surya.agustian@uin-suska.ac.id, ⁴pizaini@uin-suska.ac.id

Received May 2nd 2024; Revised Jun 10th 2024; Accepted Jun 14th 2024
Corresponding Author: Nazruddin Safaat Harahap

Abstract

Fiqh, as a vast field of study, often gives rise to various issues and differences in views among its schools of thought. Scholars' perspectives on fiqh issues aim to enrich understanding options, not to cause division. The existence of schools of thought is important for the lay Muslim community in understanding Islamic law, as it aids in interpreting the Qur'an and Hadith for everyday matters. Presently, information delivery can be quick and easy, one way being through question-and-answer applications or Question Answering Systems (QAS) related to topics users want to know about. Therefore, this research aims to create a web-based QAS about the fiqh of the four schools of thought using LangChain technology and the Large Language Model (LLM). LangChain and LLM models are capable of providing answers to questions related to Portable Document Format (PDF) files. The QAS is trained using a dataset consisting of PDF files and leverages LLM models to generate relevant text responses to user queries. The developed system successfully responds to users with testing using BERTScore, achieving an average precision of 80%, recall of 81%, and an F1 score of 81%. Meanwhile, ROUGEScore got an average value from ROUGE-1 of 56%, 58%, and 56%, ROUGE-2 of 33%, 33%, 33%, and ROUGE-L of 43%, 44%, and 43%.

Keyword: BERTScore, Four Schools of Islamic Jurisprudence, Langchain, LLM, Question Answering Systems

Abstrak

Fikih sebagai ilmu yang luas, terkadang menimbulkan beragam persoalan dan perbedaan pandangan antara madzhab-madzhabnya. Tujuan pandangan ulama tentang isu-isu fikih adalah untuk memperkaya opsi pemahaman, bukan menyebabkan perpecahan. Keberadaan mazhab penting bagi umat Islam awam dalam memahami hukum Islam, karena membantu dalam menafsirkan Al-Qur'an dan Hadits untuk masalah sehari-hari. Pengiriman informasi saat ini dapat dilakukan dengan cepat dan mudah, salah satunya melalui aplikasi tanya jawab atau *Question Answering System (QAS)* terkait materi yang ingin diketahui oleh pengguna. Sehingga pada penelitian ini bertujuan membuat sebuah QAS berbasis web tentang fikih empat madzhab menggunakan teknologi *LangChain* dan *Large Language Model (LLM)*. *LangChain* dan model LLM mampu memberikan jawaban atas pertanyaan terkait file *Portable Document Format (PDF)*. QAS dilatih menggunakan kumpulan data berupa file PDF serta memanfaatkan model LLM untuk menghasilkan respons teks yang relevan terhadap pertanyaan yang diajukan oleh pengguna. Sistem yang telah dikembangkan berhasil memberikan respons kepada pengguna dengan pengujian menggunakan *BERTScore* yang mendapatkan nilai rata-rata dari *precision* sebesar 80%, *recall* sebesar 81%, dan *f-1 score* sebesar 81%. Sedangkan *ROUGEScore* mendapatkan nilai rata-rata dari *ROUGE-1* sebesar 56%, 58%, dan 56%, *ROUGE-2* sebesar 33%, 33%, 33%, dan *ROUGE-L* sebesar 43%, 44%, dan 43%.

Kata Kunci: *BERTScore, Fikih Empat Madzhab, Langchain, LLM, Question Answering Systems*

1. PENDAHULUAN

Agama Islam dianggap sebagai agama yang komprehensif dan lengkap di antara agama-agama lainnya. Kehadirannya mencakup setiap aspek kehidupan manusia, mulai dari tahap awal kehidupan di dalam kandungan ibu hingga kembali ke alam akhirat. Ini mencakup kegiatan sehari-hari manusia seperti makan,

minum, tidur, dan interaksi dengan sesama muslim dan manusia pada umumnya. Kelebihan agama islam terletak pada kesempurnaannya, di mana hukum dan normanya dianggap lengkap, menyeluruh, dan merupakan penutup ajaran-ajaran agama Islam yang terdapat dalam ayat terakhir [1].

Dalam agama Islam, terdapat prinsip-prinsip yang terstruktur dalam ilmu-ilmu keislaman, yang membentuk suatu sistem dan mencakup hukum-hukum yang berada di bawah payungnya. Tujuannya adalah untuk memberikan kemudahan bagi para mujtahid dalam menanggapi serta menjawab permasalahan-permasalahan kontemporer. Prinsip-prinsip hukum (fikih) tidaklah disusun pada suatu periode tertentu, melainkan disusun secara terstruktur di masa yang kemudian sejalan dengan perkembangan dan pertumbuhan ijtihad di kalangan ahli dan tokoh pendiri madzhab dalam hukum Islam [2].

Fikih adalah ilmu yang luas dengan cabang-cabang tak terbatas, menimbulkan persoalan bervariasi, dan perbedaan pandangan antara madzhab-madzhab fikih dan ilmu fikih lainnya yang menciptakan ketidaksepakatan di kalangan umat muslim [3]. Pemahaman pandangan ulama tentang isu-isu fikih bertujuan untuk memperkenalkan berbagai sudut pandang guna memperkaya opsi, bukan menyebabkan perpecahan di antara umat Islam. Ulama masa lalu menekankan pentingnya kajian dalam ilmu fikih setelah memahami Al-Qur'an dan sunnah secara mendalam [4]. Mazhab merujuk pada pandangan mujtahid Islam yang berdasarkan tafsir rinci dan prinsip-prinsip. Keberadaan mazhab berpenting bagi kalangan awam dalam memahami hukum Islam, karena tidak selalu mampu langsung menafsirkan Al-Qur'an dan Hadits untuk menyelesaikan masalah sehari-hari. Namun, ada pandangan bahwa fikih hanya berdasarkan pendapat manusia tanpa dasar pada dalil syar'i, yang membuat beberapa individu tidak sepenuhnya memahami konsep ilmu fikih [5], [6].

Dengan kemajuan yang cepat dari teknologi dan informasi pada saat ini, hingga kita semakin sulit untuk terlepas dari perkembangannya. Hal ini menyebabkan arus informasi berkembang dengan cepat [7]. Pengiriman informasi saat ini dapat dilakukan dengan cepat dan mudah, salah satunya melalui aplikasi tanya jawab atau *Question Answering System (QAS)* terkait materi yang ingin diketahui oleh pengguna [8]. QAS adalah suatu sistem yang memungkinkan pengguna untuk mengajukan pertanyaan dalam bahasa sehari-hari dan mendapatkan jawaban secara cepat dan ringkas, kadang-kadang disertai dengan informasi tambahan untuk mendukung kebenaran jawaban tersebut [9]. Dalam bahasa sehari-hari atau umumnya aplikasi yang bertipe QAS ini disebut dengan *chatbot* [10].

Berdasarkan permasalahan yang telah dijelaskan di atas, maka diperlukan sebuah penyelesaiannya dengan menciptakan platform pembelajaran yang menarik minat masyarakat untuk memahami isu-isu hukum islam, terutama dalam bidang fikih, melalui pembuatan sistem tanya jawab untuk mempermudah pemahaman fikih dari empat madzhab, dengan tujuan mengurangi perbedaan pemahaman dalam praktik ibadah, khususnya dalam konteks hukum syariah [3]. Penerapan sistem tanya jawab merupakan metode paling efektif dalam mengakses informasi dari berbagai sumber pengetahuan yang tersedia untuk pengguna akhir, sehingga mereka dapat memperoleh hasil yang sesuai dengan pertanyaan yang diajukan [11].

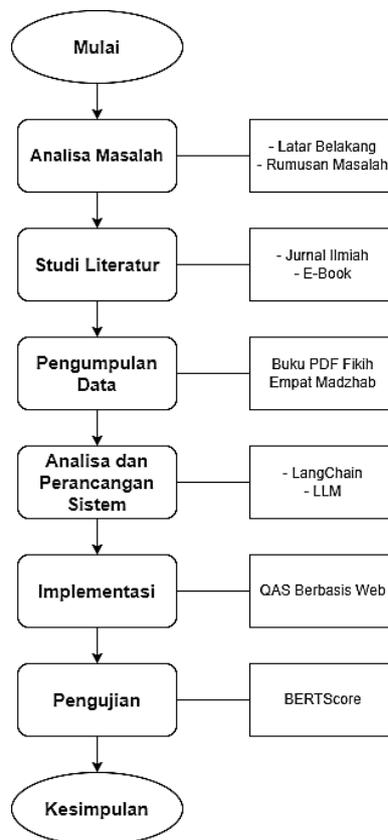
Terdapat penelitian sebelumnya mengenai aplikasi terkait. Pertama pada penelitian Fadlan Fadilah tahun 2020 yaitu mengenai aplikasi tanya jawab atau *chatbot* yang bertujuan untuk memberikan penjelasan informasi yang berkaitan dengan Universitas Islam Riau. *Chatbot* yang dibuat pada penelitian ini berbasis web yang mampu memberikan respon yang akurat terhadap pertanyaan [10]. Penelitian kedua yaitu Oguzhan Topsakal dan T. Cetin Akinci tahun 2023 tentang pemanfaatan *Large Language Model (LLM)* dalam menciptakan aplikasi dengan cepat, terutama dalam konteks *LangChain* yang memungkinkan pengembang untuk memanfaatkan LLM dengan baik, efektif dan efisien [12]. Dan pada penelitian Fatih Soygazi dan Damla Oguz tahun 2023 tentang pemanfaatan *LangChain* dan teknologi canggih LLM dalam pendidikan matematika. Pemanfaatan LLM dalam penelitian ini mampu menganalisis pertanyaan-pertanyaan matematika dapat secara signifikan meningkatkan proses pembelajaran [13].

Berdasarkan penelitian sebelumnya yang telah dijelaskan di atas, belum ada penelitian yang secara komprehensif membahas penerapan *LangChain* dalam implementasi *Question Answering System*. Oleh karena itu, peneliti melakukan penelitian tentang implementasi *Question Answering* berbasis web pada data buku fikih empat madzhab. Penelitian ini mengandalkan keunggulan *LangChain* pada aplikasi *Natural Language Model (NLP)* dan model LLM untuk mengembangkan QAS yang mampu memberikan jawaban atas pertanyaan terkait file *Portable Document Format (PDF)*. QAS dilatih menggunakan kumpulan data berupa file PDF serta memanfaatkan model LLM untuk menghasilkan respons teks yang relevan terhadap pertanyaan yang diajukan oleh pengguna [14]. *LangChain* digunakan sebagai kerangka kerja untuk mendukung QAS yang menggunakan LLM dari OpenAI sebagai penyedia API dan model bahasa. Penelitian ini bertujuan untuk membantu mencari informasi tentang fikih empat madzhab dengan cara yang akurat, efektif, dan cepat. Sehingga dengan cepat mampu memberikan jawaban secara langsung terkait dengan pertanyaan dari pengguna.

2. METODOLOGI PENELITIAN

Metodologi penelitian merupakan suatu proses yang sistematis dan teliti dalam menggali serta menelaah masalah tertentu menggunakan metode ilmiah. Tujuannya adalah mengumpulkan, memproses,

menganalisis data dengan cermat serta mengambil kesimpulan secara terstruktur dan obyektif. Semua ini dilakukan guna menyelesaikan permasalahan yang ada atau menguji hipotesis demi memperoleh pengetahuan yang bermanfaat bagi kehidupan manusia [15]. Adapun tahapan penelitian yang akan digunakan dalam menyelesaikan penelitian ini ditunjukkan pada gambar 1.



Gambar 1. Tahapan Penelitian

2.1. Analisa Masalah

Mengidentifikasi masalah dan memahami batasannya merupakan langkah pertama dalam proses penelitian. Dengan melakukan analisis ini, diharapkan peneliti mampu memahami permasalahan yang terkait dengan penelitian yang dilakukan baik dari segi fungsionalitas secara komprehensif ataupun membantu orang dalam mendalami pengetahuannya yang akan membantu dalam pemahaman yang lebih yang lebih baik [16]. Pada tahap ini, data yang digunakan terdiri dari buku PDF yang mencakup ayat-ayat Al-Qur'an, namun disadari bahwa penyalinan ayat-ayat Al-Qur'an tidak berlangsung dengan sempurna. Oleh karena itu, peneliti melakukan koreksi data dengan mengubah file ke format *txt* terlebih dahulu kemudian menyalin kembali ayat-ayat tersebut untuk memastikan keakuratan penggunaan data tersebut.

2.2. Studi Literatur

Dalam tahapan ini yang dilakukan oleh peneliti yaitu mengumpulkan semua informasi yang diperlukan dalam penelitian. Dalam hal ini, peneliti menggunakan berbagai metode untuk mengumpulkan data yang relevan sesuai dengan kebutuhan penelitian yang sedang dilakukan, di antaranya adalah jurnal ilmiah, buku, *e-book*, dan juga situs web.

2.3. Pengumpulan Data

Tujuan dari proses pengumpulan data adalah untuk memberikan informasi yang mendukung kepada peneliti dalam mengembangkan penelitian serta memperoleh informasi yang relevan dan bermanfaat untuk memahami fenomena yang sedang diteliti yaitu terkait fikih empat madzhab [17]. Pengumpulan data dilakukan dengan mencari data berupa buku pdf tentang fikih empat madzhab yang menjadi validator data tersebut adalah pakar fikih. Peneliti mengambil data yang akan dikelola untuk QAS yaitu berupa *e-book* fikih empat madzhab karya Syaikh Abdurrahman Al-Juzairi yang diakses dari link <https://drive.google.com/open?id=1kUKinqkNzmYu0vhJNbyuA5ctuHTOSiCR>. Akses *google drive* tersebut peneliti dapatkan dari sebuah grup ditelegram khusus untuk *share* tentang buku-buku islami yang diakses di <https://t.me/bukuISLAMpdf>. Untuk buku yang peneliti gunakan terdiri dari jilid 1 sampai 6. Buku fikih

empat madzhab jilid 1 terdiri dari 732 halaman, jilid 2 sebanyak 703 halaman, jilid 3 sebanyak 625 halaman, jilid 4 sebanyak 534 halaman, jilid 5 sebanyak 1154 halaman, dan jilid 6 sebanyak 851 halaman.

2.4. Analisa Sistem

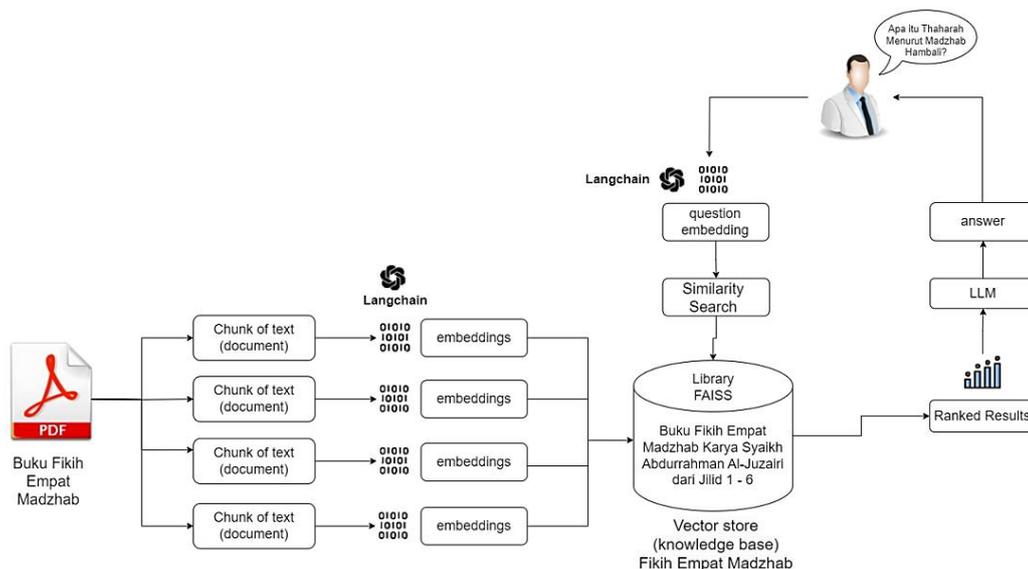
Tahapan ini untuk memenuhi kebutuhan aplikasi yang mana tahapan ini tahapan investigasi dan penentuan kebutuhan esensial. Dalam penelitian ini, analisis kebutuhan sistem menggunakan *LangChain* dan model *Large Language Model (LLM)* untuk mengembangkan *Question Answering System (QAS)* yang dapat memberikan jawaban atas pertanyaan dari file PDF serta mengevaluasi kesesuaian jawaban.

LangChain merupakan sebuah platform yang didesain untuk membangun aplikasi yang memanfaatkan model bahasa yang besar. Fokus utamanya adalah untuk memfasilitasi pengembang agar dapat dengan mudah memanfaatkan sumber data tambahan dan berinteraksi dengan aplikasi lainnya [12]. Pada penelitian ini *LangChain* berperan untuk mempermudah dalam pengembangan *Question Answering System (QAS)*.

LLM merupakan sebuah algoritma kecerdasan buatan baru yang dilatih untuk mengantisipasi urutan kata-kata tertentu berdasarkan konteks kata sebelumnya. *Generative Pre-trained Transformer (ChatGPT)* termasuk dalam kategori LLM yang dibuat oleh OpenAI dan dirilis pada bulan November tahun 2022 [18]. Pada penelitian ini LLM berperan dalam memberikan bobot nilai tertinggi dari beberapa jawaban yang diberikan oleh sistem, yang mana jawaban dengan bobot tertinggi tersebut akan dijadikan sebagai jawaban atas pertanyaan dari *user*.

2.5. Perancangan Arsitektur Sistem

Perancangan arsitektur sistem untuk QAS fikih empat madzhab menggunakan *LangChain* dan LLM. Arsitektur sistem QAS ditunjukkan pada gambar 2.



Gambar 2. Arsitektur Sistem

Tahapan awal dalam pengembangan sistem ini adalah melalui tahap *preprocessing* data, yaitu *chunk of text* dan *embeddings*. Tahapan *chunk of text* merupakan proses mengelompokkan informasi menjadi bagian-bagian yang lebih kecil sehingga mempermudah pemahaman dan proses menyajikan informasi. Dalam membagi kalimat menjadi bagian-bagian yang lebih besar, setiap bagian sesuai dengan unit sintaksis seperti frasa kata benda atau frasa kata kerja [19]. Dalam proses *chunking*, metode hanya mengekstraksi kata kunci dengan frasa terpanjang. Setelah itu, semua kata kunci yang ditemukan akan dijadikan kata kunci tanpa adanya proses seleksi atau evaluasi lebih lanjut. Dengan kata lain, tidak ada pemilahan atau penilaian lebih lanjut terhadap kata kunci yang telah diekstraksi [20]. Ukuran pemotongan teks ialah sebesar 1000 karakter perpotongannya.

Kemudian dilakukan *embedding*, yaitu proses mengkonversi kata-kata dari bahasa manusia ke dalam format yang dapat dimengerti oleh mesin yaitu bentuk *vector numerik* 0101 dan disimpan ke dalam memori *LangChain*. Dalam proses *embedding*, kata-kata yang serupa akan memiliki representasi nilai yang mendekati atau sama [21]. Setelah proses *embedding* selesai, segala potongan teks (vektor) disimpan dalam *vector stores* dengan maksud agar model *Large Language Model (LLM)* dapat mengenali keterkaitan antarkata-kata. Dalam penelitian ini, peneliti menggunakan sebuah perpustakaan (*library*) bernama FAISS untuk menyimpan *vector stores* tersebut. Sehingga untuk data dari buku pdf tersebut telah menjadi data yang

siap dipakai yang sudah dijadikan *knowledge base* yang digunakan sebagai database dari *question answering system*.

Bila ada *user* bertanya pada sistem, maka pertanyaan tersebut melalui tahapan *embedding* yang diubah menjadi representasi vektor dan kemudian dilakukan pencarian kemiripan (*similarity search*) pada *vector store* dengan bantuan *LangChain* untuk menemukan beberapa jawaban yang sesuai dengan pertanyaan tersebut. Setelah beberapa jawaban yang sesuai ditemukan, jawaban-jawaban tersebut diurutkan berdasarkan relevansinya, dan kemudian diproses oleh LLM untuk memberikan bobot tertinggi sehingga akan menghasilkan jawaban yang paling sesuai.

2.6. Implementasi

Implementasi sistem adalah langkah-langkah untuk menerapkan dan menjalankan sistem yang telah dirancang ke dalam kegiatan sehari-hari. Ini melibatkan instalasi perangkat lunak, konfigurasi, pengujian, dan memastikan bahwa sistem berfungsi dengan baik sesuai kebutuhan.

1. Pengembangan Web

Dalam pengembangan sistem yang dibuat, peneliti memanfaatkan sebuah *framework streamlit*. *Framework* ini menjadi pilihan karena kemampuannya dalam memfasilitasi pengembangan aplikasi web interaktif secara cepat dan efisien menggunakan bahasa pemrograman *Python*, tanpa memerlukan tingkat keahlian yang tinggi dalam pembuatan antarmuka pengguna yang kompleks. Dengan menggunakan *streamlit*, pengembang dapat fokus pada pengimplementasian fitur-fitur inti dari sistem yang sedang diteliti, sambil tetap memberikan pengalaman pengguna yang responsif dan mudah diakses.

2. Pengembangan API

Untuk mendukung pengembangan sistem, peneliti memerlukan akses ke sebuah *Application Programming Interface* (API). API merupakan kumpulan metode perangkat lunak yang memungkinkan komunikasi antara dua aplikasi melalui serangkaian aturan yang telah ditentukan oleh penyedia layanan tersebut. Peneliti memilih menggunakan layanan OpenAI untuk mengakses API. Key API ini akan digunakan untuk mengotentikasi setiap permintaan yang dikirim ke server OpenAI agar dapat menggunakan model bahasa LLM, yakni *Generative Pre-trained Transformer* (GPT). Dalam penelitian ini, peneliti menggunakan versi GPT-3.5 Turbo.

2.7. Pengujian

Tujuan pengujian ini untuk menilai apakah sistem yang telah dibuat dapat digunakan secara efektif atau tidak. Dalam penelitian ini, pengujian dilakukan dengan menguji apakah jawaban yang dihasilkan oleh sistem sesuai dengan data referensi yang telah ada, yaitu dengan menggunakan metode *Bidirectional Encoder Representations from Transformers* (BERT) dan *Recall-Oriented Understudy for Gisting Evaluation* (ROUGE).

1. BERTScore

BERTScore merupakan sebuah model deep learning yang telah memberikan kemajuan yang signifikan dalam berbagai tugas *Natural Language Processing* (NLP) [22]. BERT dirancang untuk memahami kalimat ambigu dengan menggunakan konteks teks sekitarnya, membangun pemahaman yang lebih lengkap melalui *transformer*. *Transformer*, dengan mekanisme *self-attention*, belajar dan menyesuaikan pemahaman terhadap hubungan kontekstual antar kata [23]. BERT memanfaatkan *encoder* dalam *transformator* sebagai komponen dalam model *pre-training* untuk berbagai tugas pemrosesan bahasa alami seperti *Sentiment Analysis* (SA), *Question Answering* (QA), dan *Text Summarization* (TS). Secara praktis, BERT melalui dua tahap dalam prosesnya, yakni *pre-training* untuk memahami bahasa dan *fine-tuning* untuk tugas-tugas spesifik. Berikut adalah persamaan untuk menghitung *BERTScore* [24].

a. Precision

Precision adalah ukuran dari sejauh mana prediksi model sesuai dengan data yang diminta.

$$P_{\text{BERT}} = \frac{1}{|\hat{x}|} \sum_{\hat{x}_j \in \hat{x}} \max x_i^T \hat{x}_j \quad (1)$$

b. Recall

Recall mengukur seberapa efektif model dalam memprediksi kelas positif secara benar.

$$R_{\text{BERT}} = \frac{1}{x} \sum_{x_j \in x} \max x_i^T \hat{x}_j \quad (2)$$

c. *F-1 Score*

F1-Score adalah nilai rata-rata yang membandingkan *precision* dan *recall*.

$$F_{BERT} = 2 \frac{P_{BERT} \cdot R_{BERT}}{P_{BERT} + R_{BERT}} \quad (3)$$

2. *ROUGEScore*

ROUGEScore digunakan untuk mengukur kualitas ringkasan otomatis dengan membandingkannya antara ringkasan yang dihasilkan oleh sistem dengan ringkasan ideal yang dibuat oleh manusia. Nilai ROUGE diperoleh dengan menghitung jumlah kata yang sama, seperti n-gram, urutan kata, dan pasangan kata [25]. Pemilihan ROUGE dalam evaluasi ringkasan dilakukan karena metrik ini telah menjadi standar evaluasi yang umum dan banyak digunakan dalam penelitian sebelumnya [26]. *ROUGEScore* terdiri dari *precision*, *recall*, dan *f1-score* untuk ROUGE-N, dengan nilai N adalah 1, 2, atau L. Variabel N menunjukkan jumlah kata berurutan yang sama dalam kedua dokumen yang dibandingkan, yaitu ringkasan metode MMR dan ringkasan manusia (*gold standard*). N bernilai 1 menghitung unigram (*word-1-gram*) dan N bernilai 2 menghitung bigram (*word-2-gram*) dalam kedua dokumen. Sedangkan variabel N yang bernilai L adalah *Longest Common Subsequence* (LCS), yaitu rangkaian kata terpanjang yang muncul secara berurutan dalam kedua dokumen yang dibandingkan. Berikut adalah persamaan untuk menghitung *ROUGEScore* [27].

a. *Precision*

Precision adalah metode untuk mengukur jumlah prediksi yang relevan dengan menghitung jumlah kata yang sama, baik itu *unigram*, *bigram*, atau LCS, lalu membaginya dengan total kata dalam ringkasan yang dihasilkan oleh sistem.

$$ROUGE - 1 \text{ precision} = \frac{\text{jumlah unigram kata sama}}{\text{keseluruhan kata diringkasan sistem}} \quad (4)$$

$$ROUGE - 2 \text{ precision} = \frac{\text{jumlah bigram kata sama}}{\text{keseluruhan kata diringkasan sistem}} \quad (5)$$

$$ROUGE - L \text{ precision} = \frac{\text{Longest Common Subsequence (LCS)}}{\text{keseluruhan kata diringkasan sistem}} \quad (6)$$

b. *Recall*

Recall adalah metode untuk mengukur jumlah prediksi yang relevan dengan menghitung jumlah kata yang sama, baik *unigram*, *bigram*, atau LCS, kemudian membaginya dengan total kata dalam ringkasan manusia.

$$ROUGE - 1 \text{ recall} = \frac{\text{jumlah unigram kata sama}}{\text{keseluruhan kata diringkasan manusia}} \quad (7)$$

$$ROUGE - 2 \text{ recall} = \frac{\text{jumlah bigram kata sama}}{\text{keseluruhan kata diringkasan manusia}} \quad (8)$$

$$ROUGE - L \text{ recall} = \frac{\text{Longest Common Subsequence (LCS)}}{\text{keseluruhan kata diringkasan manusia}} \quad (9)$$

c. *F-1 Score*

F1-Score adalah metode untuk mengukur rata-rata *harmonis* (*harmonic mean*) antara *recall* dan *precision*.

$$F-1 \text{ score} = 2 \times \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$

3. HASIL DAN PEMBAHASAN

3.1. Implementasi

Hasil dari implementasi *Question Answering System* fikih empat madzhab berbasis web ditunjukkan pada gambar 3.

Tampilan pada gambar 3 merupakan antarmuka untuk memulai berinteraksi dengan sistem tanya jawab fikih empat madzhab. Tampilan sistem ini dibuat menggunakan platform *streamlit*. *Question Answering System* (QAS) ini dapat diakses oleh siapa pun yang ingin memperoleh informasi tentang fikih empat madzhab. Pengguna hanya dapat mengajukan pertanyaan seputar fikih empat madzhab dan sistem

akan memberikan jawaban berdasarkan informasi yang terdapat dalam enam buku fikih empat madzhab yang telah diprogramkan. Pada tampilan tersebut, pengguna mengajukan pertanyaan tentang "apa pengertian thaharah menurut madzhab Hambali," dan sistem memberikan jawaban mengenai pengertian dari thaharah berdasarkan informasi yang ada dalam buku-buku fikih empat madzhab yang telah diprogramkan. Selain itu, sistem juga menampilkan empat referensi yang dikutip dari buku-buku fikih empat madzhab tersebut.



Gambar 3. Tampilan Sistem

3.2. Pengujian

Dalam penelitian ini, pengujian evaluasi jawaban sistem melalui pengujian validasi jawaban sistem dan pengujian untuk menilai akurasi jawaban dari sistem. Pengujian dilakukan menggunakan *BERTScore* dan juga *ROUGEScore* menggunakan 10 sampel pertanyaan. Pertanyaan yang diajukan pada sistem ini merupakan pertanyaan dari beberapa individu yang sepenuhnya kurang memahami kompleksitas atau detail dari topik yang mereka tanyakan. Sampel pertanyaan ditunjukkan pada tabel 1.

Tabel 1. Sampel Pertanyaan

No	Pertanyaan
1	Apa pengertian thaharah menurut madzhab Hambali?
2	Apa pengertian muzaraah menurut madzhab Hanafi?
3	Apa saja yang dapat membatalkan wudhu?
4	Apa hukum seorang istri yang keluar rumah tanpa seizin suaminya?
5	Pada perihal tayamum, apakah bisa tayamum di dalam kendaraan umum untuk melakukan sholat sedangkan kita mengetahui jarak yang kita tempuh sudah dekat?
6	Pada kasus makmum laki-laki yang sejajar dengan salah satu dari sebagian makmum perempuan, apakah dia harus mengulang shalat dan apakah shalatnya sah?
7	Siapa yang berhak menjadi wali nikah bagi pengantin perempuan yang ayahnya telah meninggal?
8	Menurut madzhab Maliki apa yang dimaksud air suci tetapi tidak mensucikan?
9	Menurut madzhab Mailiki apakah boleh seorang istri menjatuhkan talak kepada suami yang tidak pernah memberi nafkah istrinya?
10	Menurut madzhab Syafi'i apa hukuman orang yang melakukan pembunuhan?

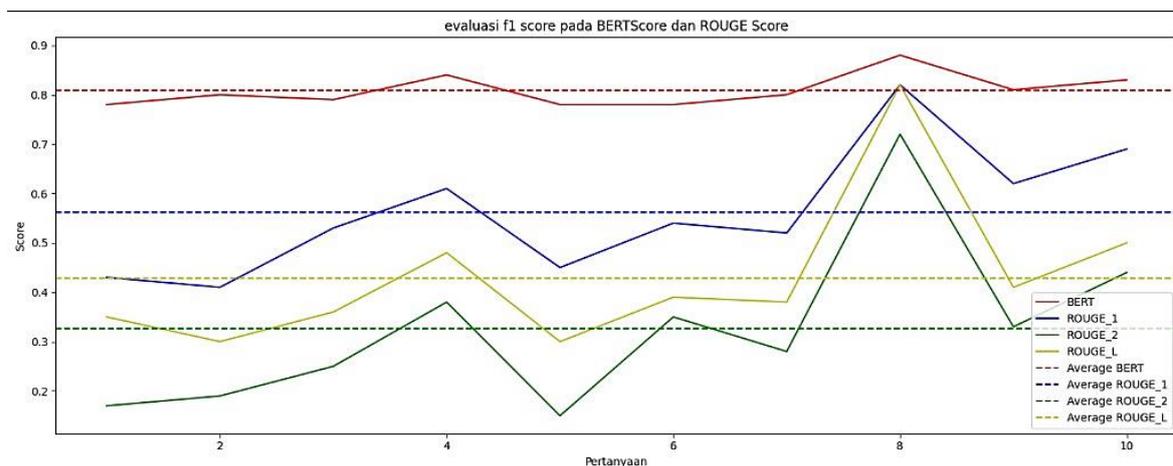
Pengujian evaluasi *Question Answering System* (QAS) dengan menggunakan *BERTScore* melibatkan penilaian seberapa baik jawaban yang dihasilkan oleh sistem tersebut cocok dengan jawaban referensi. *BERTScore* mengukur kesamaan makna antara jawaban sistem dan jawaban referensi dengan membandingkan representasi vektor kata yang dihasilkan dari model BERT. Skor yang lebih tinggi menunjukkan bahwa jawaban sistem lebih mirip secara makna dengan jawaban referensi, yang menunjukkan bahwa QAS memiliki kinerja yang lebih baik dalam memahami dan menjawab pertanyaan secara benar. Sedangkan pengujian ROUGE untuk mengevaluasi berbasis teks dan lebih mementingkan kesamaan literal antara jawaban yang dihasilkan oleh sistem dan jawaban referensi.

Tabel 2 menunjukkan hasil *precision*, *recall* dan *f-1 score* dari pengujian menggunakan *BERTScore* dan *ROUGEScore*.

Tabel 2. Hasil Pengujian BERTScore dan ROUGEScore

Pertanyaan	BERTScore(%)			ROUGE-1(%)			ROUGE-2(%)			ROUGE-L(%)		
	P	R	F-1	P	R	F-1	P	R	F-1	P	R	F-1
1	78%	77%	77%	36%	54%	43%	14%	21%	17%	29%	44%	35%
2	79%	81%	80%	38%	45%	41%	17%	21%	19%	28%	33%	30%
3	78%	79%	79%	54%	52%	53%	25%	24%	25%	37%	36%	36%
4	84%	84%	84%	62%	60%	61%	38%	37%	38%	49%	48%	48%
5	80%	76%	78%	39%	54%	45%	13%	18%	15%	26%	36%	30%
6	78%	78%	78%	59%	51%	54%	38%	33%	35%	42%	37%	39%
7	78%	81%	80%	53%	51%	52%	29%	28%	28%	39%	37%	38%
8	86%	90%	88%	87%	77%	82%	77%	68%	72%	87%	77%	82%
9	80%	82%	81%	69%	56%	62%	37%	30%	33%	46%	37%	41%
10	84%	82%	83%	63%	75%	69%	41%	48%	44%	46%	54%	50%
Rata-rata	80%	81%	81%	56%	58%	56%	33%	33%	33%	43%	44%	43%

Hasil pengujian menggunakan BERTScore mendapatkan nilai rata-rata dari *precision* sebesar 80%, *recall* sebesar 81%, dan *f-1 score* sebesar 81%. Sedangkan untuk pengujian menggunakan ROUGEScore mendapatkan nilai rata-rata dari ROUGE-1 untuk nilai *precision* sebesar 56%, *recall* sebesar 58%, dan *f-1 score* sebesar 56%, ROUGE-2 sebesar 33%, 33%, 33%, dan untuk ROUGE-L sebesar 43%, 44%, dan 43%. Berdasarkan hasil pengujian yang telah dilakukan dengan kedua pengujian tersebut, maka dapat disimpulkan bahwa BERTScore menunjukkan skor yang lebih tinggi dibandingkan dengan ROUGEScore, karena ROUGE lebih efektif digunakan untuk tugas peringkasan dokumen. Gambar 4 menunjukkan hasil pengujian BERTScore dan ROUGEScore.



Gambar 4. Hasil Pengujian BERTScore dan ROUGEScore

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, penulis berhasil membuat sebuah *Question Answering System* (QAS) berbasis web dengan menerapkan *LangChain* dan *Large Language Model* (LLM) menggunakan data berupa file *Portable Document Format* (PDF) dari buku fikih empat madzhab karya Syaikh Abdurrahman Al-Juzairi dari jilid 1 sampai 6. Pengujian yang dilakukan terhadap sistem menggunakan metode BERTScore mendapatkan nilai rata-rata dari *precision* sebesar 80%, *recall* sebesar 81%, dan *f-1 score* sebesar 81% dan menggunakan metode ROUGEScore mendapatkan nilai rata-rata dari ROUGE-1 untuk nilai *precision* sebesar 56%, *recall* sebesar 58%, dan *f-1 score* sebesar 56%, ROUGE-2 sebesar 33%, 33%, 33%, dan untuk ROUGE-L sebesar 43%, 44%, dan 43% yang dapat disimpulkan bahwa implementasi *Question Answering System* yang di lakukan memiliki tingkat akurasi yang cukup tinggi dalam mengatasi pertanyaan terkait data fikih empat madzhab.

Hal ini menunjukkan bahwa sistem sudah mampu memberikan jawaban yang relevan dan memuaskan bagi pengguna, serta memudahkan pemahaman konsep-konsep dalam fikih empat madzhab. Namun, penelitian ini juga menyoroti beberapa area yang perlu ditingkatkan di masa depan. Salah satunya adalah meningkatkan kinerja dan fungsionalitas sistem, mungkin dengan mengintegrasikan teknologi baru seperti kecerdasan buatan yang lebih canggih atau analisis data yang lebih mendalam. Saran untuk penelitian selanjutnya adalah memperbaiki data dan konsistensi dalam penyajian hasil jawaban agar lebih sesuai dengan pertanyaan pengguna.

REFERENCES

- [1] M. D. Faza, "Tinjauan Fiqih Empat Mazhab Terhadap Acara Walimah," *Al Fuadiy (Jurnal Huk. Kel. Islam.*, vol. 4, no. 2, pp. 17–25, 2022.
- [2] S. R. F. Sanusi, "Kaidah Fikih: Sejarah dan Pemikiran Empat Madzhab," *Tahkim J. Perad. dan Huk. Islam.*, vol. 4, no. 2, pp. 23–46, Nov. 2021, doi: 10.29313/tahkim.v4i2.6809.
- [3] S. Rahayu and S. Ma'mun, "Rancang Bangun Aplikasi Fiqih Ibadah 4 Madzhab Berbasis Android," *J. Algoritma.*, vol. 18, no. 1, pp. 41–49, Aug. 2021, doi: 10.33364/algoritma/v.18-1.833.
- [4] S. A. Al-Juzairi, *Fiqih Empat Madzhab*, Jilid 1. PUSTAKA AL-KAUTSAR, 2008.
- [5] A. Faqih, "Madzhab Google : Peran Vital Guru Fiqih di Era Disrupsi," *Dar El Ilmi J. Keagamaan, Pendidik. dan Hum.*, vol. 10, no. 1, pp. 120–134, 2023, doi: <https://doi.org/10.52166/darelilmi.v10i1.4728>.
- [6] M. Ajib, *Masalah Khilafiyah 4 Madzhab Terpopuler*. Jakarta Selatan: Rumah Fiqih Publishing, 2018.
- [7] E. Rahmawati and E. Abdulmanan, "Pemodelan Aplikasi Dunia Islam Mengaji Berbasis Android," *J. Ilmu Pengetah. dan Teknol. Komput.*, vol. 4, no. 2, pp. 197–20, 2019, doi: <https://doi.org/10.33480/jitk.v4i2>.
- [8] D. Apriliani, S. F. Handayani, T. N. Anugrahaeni, A. Miftahudin, L. Nurarifiah, and I. T. Saputra, "Aplikasi Question Answer Sebagai Media Pembelajaran Interaktif Untuk Mata Pelajaran Akuntansi," *JMM J. Masy. Mandiri.*, vol. 7, no. 2, p. 2003, 2023, doi: <https://doi.org/10.31764/jmm.v7i2.13867>.
- [9] A. Setiawan, O. N. Pratiwi, and R. Y. Fa'rifah, "Question Answering System Dalam Bentuk Chatbot Pada Platform Line Untuk Mata Pelajaran Sejarah SMA/MA Dengan Menggunakan Algoritma Levenshtein Distance," *eProceedings Eng.*, vol. 8, no. 5, pp. 9794–9802, 2021.
- [10] F. Fadilah, "Aplikasi Tanya Jawab Mengenai Universitas Islam Riau Berbasis Web Menggunakan Dialogflow," Universitas Islam Riau, 2020.
- [11] A. Dhandapani and V. Vadivel, "Question Answering System over Semantic Web," *IEEE Access*, vol. 9, pp. 46900–46910, 2021, doi: 10.1109/ACCESS.2021.3067942.
- [12] O. Topsakal and T. C. Akinci, "Creating Large Language Model Applications Utilizing LangChain: A Primer on Developing LLM Apps Fast," *Int. Conf. Appl. Eng. Nat. Sci.*, vol. 1, no. 1, pp. 1050–1056, Jul. 2023, doi: 10.59287/icaens.1127.
- [13] F. Soygazi and D. Oguz, "An Analysis of Large Language Models and LangChain in Mathematics Education," no. October, 2023, doi: <https://doi.org/10.1145/3633598.3633614>.
- [14] A. Pesaru, T. S. Gill, and A. R. Tangella, "AI assistant for Document Management Using Lang Chain and Pinecone," *Int. Res. J. Mod. Eng. Technol. Sci.*, no. 06, pp. 3980–3983, Jun. 2023, doi: 10.56726/IRJMETS42630.
- [15] R. Abubakar, *Pengantar Metodologi Penelitian*, 1st ed. Yogyakarta: SUKA-Press UIN Sunan Kalijaga, 2021.
- [16] A. Rahmadsyah, H. Hartono, and R. Rosnelly, "Analisa Association Rule Pada Algoritma Apriori Untuk Minat Pembelian Alat Kesehatan," *J. Media Inform. Budidarma*, vol. 5, no. 1, p. 280, Jan. 2021, doi: 10.30865/mib.v5i1.2658.
- [17] L. Nur Aziza, R. Yuli Astuti, B. Akbar Maulana, and N. Hidayati, "Application of the K-Nearest Neighbor Algorithm for Food Security Classification in Central Java Province," *MALCOM Indones. J. Mach. Learn. Comput. Sci. J.*, vol. 4, no. 2, pp. 404–412, 2024, doi: <https://doi.org/10.57152/malcom.v4i2.1201>.
- [18] H. Hidayatullah, Purwantoro, and Y. Umaidah, "Penerapan Naïve Bayes Dengan Optimasi Information Gain Dan Smote Untuk Analisis Sentimen Pengguna Aplikasi Chatgpt," *JATI J. Mhs. Tek. Inform.*, vol. 7, no. 3, 2023, doi: <https://doi.org/10.36040/jati.v7i3.6887>.
- [19] P. R. Togatorop, R. P. Simanjuntak, S. B. Manurung, and M. C. Silalahi, "Pembangkit Entity Relationship Diagram Dari Spesifikasi Kebutuhan Menggunakan Natural Language Processing Untuk Bahasa Indonesia," *J. Komput. dan Inform.*, vol. 9, no. 2, pp. 196–206, 2021, doi: 10.35508/jicon.v9i2.5051.
- [20] I. N. P. Trisna and A. Nurwidyantoro, "Single Document Keywords Extraction in Bahasa Indonesia using Phrase Chunking," *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 18, no. 4, pp. 1917–1925, 2020, doi: 10.12928/TELKOMNIKA.V18I4.14389.
- [21] F. Fatharani, K. P. Kania, J. Hutahaean, and S. R. Wulan, "Deteksi Intensi Chatbot Berbahasa Indonesia dengan Menggunakan Metode Capsule Network," *J. Inf. Syst. Res.*, vol. 3, no. 4, pp. 590–596, Jul. 2022, doi: 10.47065/josh.v3i4.1821.
- [22] R. M. R. W. P. K. Atmaja and W. Yustanti, "Analisis Sentimen Customer Review Aplikasi RuangGuru dengan Metode BERT (Bidirectional EncoderRepresentations from Transformers)," *JEISBI (Journal Emerg. Inf. Syst. Bus. Intell.*, vol. 02, no. 3, pp. 55–62, 2021.
- [23] W. Alan, "Klasifikasi Emosi pada Teks Menggunakan Deep Learning," 2021.
- [24] F. Fajri, B. Tutuko, and S. Sukemi, "Membandingkan Nilai Akurasi BERT dan DistilBERT pada Dataset Twitter," *JUSIFO (Jurnal Sist. Informasi)*, vol. 8, no. 2, pp. 71–80, 2022, doi:

- 10.19109/jusifo.v8i2.13885.
- [25] Y. Musa'adah, Y. Wibisono, and Y. Wihardi, "Deep Reinforced Model dan Rules-Based untuk Peringkasan Kalimat Bahasa Indonesia," *JLK (Jurnal Linguist. Komputasional)*, vol. 3, no. 2, p. 40154, 2020, doi: <https://doi.org/10.26418/jlk.v3i2.34>.
- [26] Nyoman Purnama and Ni Nengah Widya Utami, "Implementasi Peringkas Dokumen Berbahasa Indonesia Menggunakan Metode Text to Text Transfer Transformer (T5)," *J. Teknol. Inf. dan Komput.*, vol. 9, no. 4, pp. 381–391, 2023.
- [27] Z. Idhafi, S. Agustian, F. Yanto, and N. Safaat H, "Peringkas Teks Otomatis pada Artikel Berbahasa Indonesia Menggunakan Metode Maximum Marginal Relevance," *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 4, no. 3, pp. 609–618, Dec. 2023, doi: [10.37859/coscitech.v4i3.6311](https://doi.org/10.37859/coscitech.v4i3.6311).