Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 293-304
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6182

Aplikasi Web Question Answering Menggunakan Langchain OpenAI Tentang Peraturan Perundang-Undangan Bidang Pendidikan

Ikhsan Dwi Saputra, Nazruddin Safaat Harahap*, Surya Agustian, Muhammad Fikry, Lola Oktavia

Fakultas Sains Dan Teknologi, Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia
Email: [1]12050113500@students.uin-suska.ac.id, [2,*]nazruddin.safaat@uin-suska.ac.id, [3]surya.agustian@uin-suska.ac.id, [4]muhammad.fikry@uin-suska.ac.id, [5]lola.oktavia@uin-suska.ac.id
Email Penulis Korespondensi: nazruddin.safaat@uin-suska.ac.id
Submitted: 01/11/2024; Accepted: 30/11/2024; Published: 30/11/2024

Abstrak−Dalam perkembangan teknologi informasi yang pesat selama beberapa tahun terakhir, kemudahan dalam mengakses informasi telah menjadi salah satu pencapaian yang signifikan. Kecerdasan buatan (AI) telah muncul sebagai alat yang potensial dalam menghadirkan solusi inovatif di berbagai sektor kehidupan manusia. Penelitian ini bertujuan untuk mengembangkan sebuah aplikasi web yang mampu menjawab pertanyaan terkait peraturan perundang-undangan di bidang pendidikan dengan memanfaatkan framework LangChain dan model BERT. Masalah utama yang dihadapi adalah kompleksitas dan volume dokumen legal yang sulit diakses dan dipahami oleh pengguna awam. Metode yang digunakan dalam penelitian ini melibatkan konversi dokumen legal dari format PDF ke teks, pemotongan teks menggunakan LangChain, dan evaluasi performa sistem menggunakan BERTScore dan ROUGE Score. Hasil penelitian menunjukkan bahwa BERTScore lebih unggul dalam mengukur kesesuaian antara jawaban sistem dan referensi, dengan beberapa pertanyaan mendapatkan skor 100%. Namun, terdapat beberapa keterbatasan seperti proses konversi dokumen yang memerlukan upaya manual dan kebutuhan sumber daya komputasi yang besar untuk pemrosesan teks. Penelitian ini memberikan kontribusi signifikan dalam memfasilitasi akses dan pemahaman dokumen legal pendidikan, serta membuka peluang untuk pengembangan lebih lanjut dengan teknik konversi dan model AI yang lebih canggih.

Kata Kunci: Question Answering; LangChain; BERTScore; ROUGE Score; Dokumen Legal; Pendidikan

Abstract−In the rapid development of information technology over the past few years, the ease of accessing information has been one of the significant achievements. Artificial intelligence (AI) has emerged as a potential tool in bringing innovative solutions in various sectors of human life. This research aims to develop a web application capable of answering questions related to educational legislation using the LangChain framework and BERT model. The primary issue addressed is the complexity and volume of legal documents that are challenging for lay users to access and understand. The methodology involves converting legal documents from PDF to text, segmenting the text using LangChain, and evaluating system performance with BERTScore and ROUGE Score. The results indicate that BERTScore is superior in measuring the alignment between the system's answers and reference answers, with some questions achieving a score of 100%. However, there are limitations, such as the manual effort required for document conversion and the substantial computational resources needed for text processing. This research significantly contributes to facilitating access and comprehension of educational legal documents and opens opportunities for further development with more advanced conversion techniques and AI models.

Keywords: Question Answering; LangChain; BERTScore; ROUGE Score; Legal Documents; Education

1. PENDAHULUAN

Dalam perkembangan teknologi informasi yang pesat selama beberapa tahun terakhir, kemudahan dalam mengakses informasi telah menjadi salah satu pencapaian yang signifikan. Kecerdasan buatan (AI) telah muncul sebagai alat yang potensial dalam menghadirkan solusi inovatif di berbagai sektor kehidupan manusia [1]. Tantangan utama yang dihadapi dalam penelitian ini adalah kompleksitas dalam mengakses dan memahami dokumen legal di bidang pendidikan. Dokumen tersebut seringkali sangat teknis dan padat dengan terminologi hukum, yang membuatnya sulit dipahami oleh orang awam. Selain itu, tingginya volume dokumen legal dan perubahan peraturan yang terjadi secara berkala membuat pencarian dan interpretasi informasi menjadi tantangan tersendiri. Kondisi ini mengakibatkan hambatan dalam mengakses informasi yang relevan dengan cepat dan tepat. Teknologi *artificial intelligence* (AI) yang terus berkembang, menawarkan berbagai peluang untuk meningkatkan efisiensi, keakuratan, dan aksesibilitas informasi. Salah satu teknologi dalam perkembangan *artificial intelligence* (AI) adalah kehadiran *GPT (Generative Prompt Transformer)* yang dirancang oleh OpenAI, yang dikenal karena kemampuannya dalam pemrosesan bahasa alami yang luar biasa [2]. Penelitian ini mengidentifikasi potensi pemanfaatan teknologi *artificial intelligence* (AI), khususnya *GPT*, untuk mengatasi kesulitan dalam mengakses dokumen legal bidang pendidikan. Dengan kemampuan pemrosesan bahasa alami yang luar biasa, model GPT menawarkan solusi inovatif untuk menafsirkan dan mengekstrak informasi hukum dari dokumen tersebut [3].

Penelitian terkait *questions anwering* telah mendalami pemanfaatan model *GPT* dengan penerapan kerangka kerja LangChain [4]. Sistem penjawaban pertanyaan telah menarik perhatian yang signifikan dalam beberapa tahun terakhir karena potensinya untuk mengubah cara kita berinteraksi dengan informasi. Kemampuan untuk mengekstrak informasi relevan dari sejumlah besar data yang tidak terstruktur memiliki banyak aplikasi [5]. Terkait model *gpt* penelitian Reiichiro Nakano, dkk. Model terbaik diperoleh dengan melakukan *fine-tuning GPT-3* menggunakan kloning perilaku, kemudian melakukan sampling penolakan terhadap model penghargaan

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 9, No. 9, Month 2022, Page 999–999
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v9i9.9999

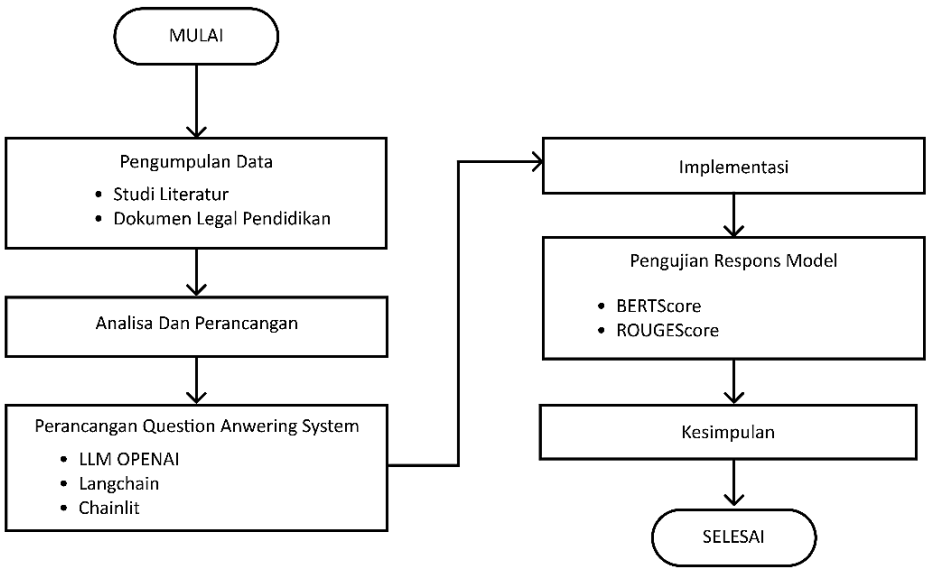
yang dilatih untuk memprediksi preferensi manusia [6]. Selanjutnya penelitian Tiago Lubiana, dkk bertujuan untuk memberikan panduan praktis dan berguna bagi para peneliti dan profesional dalam bidang biologi komputasi tentang bagaimana memanfaatkan ChatGPT secara optimal [7]. Penilitian ini juga melakukan pendekatan seperti yang dilakukan oleh Arjun Pesaru, dkk yakni Pendekatan memanfaatkan kekuatan LangChain dan Model *Large Language Models* (LLM) untuk membuat chatbot yang dapat menjawab pertanyaan tentang berkas PDF [8]. Selanjutnya penilitian ini juga memperhatikan aspek model gpt [9]. Kinerja dalam *prompting* juga di lakukan oleh Takeshi Kojima, dkk [10].

Pada pengembangan *questions answering* menggunakan model waterfall [11]. Pemahaman terhadap *natural language processing* (nlp) yang dilakukan oleh Diksha Khurana, dkk juga perlu menjadi hal yang penting dalam penilitian ini [12]. Selanjutnya model waterfall terdiri dari serangkaian tahap yang mengikuti urutan linier, dari awal hingga akhir [13]. Model ChatGpt adalah solusi yang menggabungkan Algoritma Pembelajaran Penguatan *(Reinforcement Learning)* dengan masukan manusia, dengan lebih dari 150 miliar parameter yang diintegrasikan ke dalam model [14]. Penelitian yang dilakukan oleh Gobinda G. Chowdhury, dkk juga mendalami terkait topik dari nlp, Para peneliti *Natural Language Processing* (NLP) berupaya mengumpulkan pengetahuan tentang bagaimana manusia memahami dan menggunakan bahasa sehingga alat dan teknik yang sesuai dapat dikembangkan untuk membuat sistem komputer memahami dan memanipulasi bahasa alami untuk melaksanakan tugas yang diinginkan. Penelitian ini juga memberikan dasar dasar dalam nlp terletak di sejumlah disiplin ilmu, seperti ilmu komputer dan informasi, linguistik, matematika, teknik listrik dan elektronik, kecerdasan buatan dan robotika, psikologi, dll [15]. Selanjutnya pengujian pada penilitian ini akan menggunakan bert score, penilitian yang dilakukan oleh Jacob Devlin, dkk melakukan pendekatan berbasis fine-tuning dengan mengusulkan BERT: *Bidirectional Encoder Representations from Transformers.* Menurutnya Terdapat dua strategi yang ada untuk menerapkan representasi bahasa yang telah dilatih sebelumnya ke tugas-tugas lanjutan (downstream tasks): berbasis fitur dan *fine-tuning* [16].

Penelitian terbaru telah menunjukkan peningkatan signifikan dalam banyak tugas NLP dan tolok ukur dengan melakukan pra-pelatihan pada korpus teks yang besar diikuti dengan *fine-tuning* pada tugas tertentu. Di sini, kami menunjukkan bahwa meningkatkan skala model bahasa secara signifikan meningkatkan kinerja few-shot yang tidak bergantung tugas, terkadang bahkan mencapai daya saing dengan pendekatan fine-tuning yang sebelumnya menjadi yang terbaik. Secara khusus, kami melatih GPT-3, sebuah model bahasa autoregresif dengan 175 miliar parameter, 10 kali lebih besar dari model bahasa *non-sparse* sebelumnya, dan menguji kinerjanya dalam pengaturan few-shot [17]. Penelitian yang dilakukan oleh Zeynep Akkalyoncu Yilmaz, dkk melakukan Demonstrasi yang berfokus pada tantangan teknis dalam integrasi kemampuan NLP dan IR, serta rasional desain di balik pendekatan kami untuk integrasi yang erat antara Python (untuk mendukung jaringan saraf) dan Java Virtual Machine (untuk mendukung pengambilan dokumen menggunakan pustaka pencarian open-source Lucene) [18]. Dalam memprediksi jawaban terbaik penelitian Wataru Sakata, dkk pada penilitiannya juga memanfaatkan bert score [19]. Dan diperlukan pemahaman terkait pemeringkatan dalam bert yang dilakukan dalam penelitian oleh Yifan Qiao, dkk [20].

2. METODOLOGI PENELITIAN

Proses yang diikuti dalam penelitian ini untuk merancang dan mengembangkan *questions answering* berbasis web dengan sumber konteks dokumen hukum di bidang pendidikan disajikan pada Gambar 1 berikut.



Gambar 1. Diagram Alur Penelitian

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 9, No. 9, Month 2022, Page 999–999
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v9i9.9999

Dari diagram alur penelitian yang ditampilkan pada Gambar 1, tahapan awal mencakup pengumpulan data serta perancangan sistem/metode. Tahapan ini bertujuan untuk memastikan ketersediaan informasi yang relevan dan mendukung proses penelitian sekaligus menyusun kerangka kerja yang sistematis untuk mencapai tujuan penelitian.

2.1 Pengumpulan Data

Dalam tahap awal pengumpulan data penelitian ini, fokus akan tertuju pada dua aspek penting, yaitu sub bab studi literatur yang merinci temuan penelitian terdahulu, dan sub bab dokumen legal pendidikan yang mencakup informasi relevan dan berkaitan dengan topik penelitian.

a. Studi Literatur
   Pada sub-bab studi literatur ini, penelitian akan mengeksplorasi dan menganalisis kajian-kajian terdahulu yang relevan dengan pengembangan sistem tanya jawab hukum berbasis web, terutama yang berfokus pada integrasi antara LangChain dan OpenAI. Selain itu, penelitian akan mendalami aspek dokumen hukum di bidang pendidikan. Analisis ini bertujuan untuk memahami kontribusi, metodologi, dan temuan penelitian sebelumnya yang dapat memberikan wawasan mendalam terkait implementasi teknologi ini.
b. Dokumen Legal Pendidikan
   Peneliti akan mengeksplorasi dan menguraikan data yang akan menjadi sumber konteks pada *questions answering* berbasis web ini, yaitu dokumen legal pendidikan.

2.2 Analisa Dan Perancangan

Pada tahapan analisis dan perancangan, sub bab ini akan terbagi menjadi dua aspek:

a. Analisa Data
   Pada tahap analisa data, penelitian akan meninjau dan mengevaluasi berbagai dokumen legal yang berkaitan dengan sektor pendidikan, seperti undang-undang, peraturan, kebijakan, dan pedoman. Analisis ini bertujuan untuk memberikan pemahaman mendalam tentang kerangka hukum yang mengatur pendidikan, sehingga dapat mengidentifikasi tema-tema utama dan konsep-konsep yang dapat menginformasikan pengembangan sistem tanya jawab hukum berbasis web.
b. Perancangan *Questions Answering*
   Sub-bab ini akan mendeskripsikan perancangan sistem tanya jawab menggunakan kerangka kerja LangChain. Penggunaan LangChain memerlukan perhatian terhadap beberapa tahapan penting yang harus diikuti secara sistematis. Chainlit juga akan digunakan sebagai tampilan ui dari sistem tanya jawab ini. Chainlit adalah paket Python sumber terbuka yang digunakan untuk membangun AI Percakapan siap produksi. Chainlit kompatibel dengan semua program dan pustaka Python. Meskipun demikian, Chainlit dilengkapi dengan serangkaian integrasi dengan pustaka dan kerangka kerja populer.

2.3 Implementasi

Pada tahap ke-3 dari gambar 1 yakni implementasi QAS web, penelitian ini akan menerapkan desain yang telah disusun ke dalam bentuk sistem *Questions Answering* berbasis web. Proses implementasi akan melibatkan integrasi model GPT dalam kerangka kerja LangChain, sehingga dapat memberikan layanan tanya jawab hukum secara interaktif dan responsif. Dalam pengembangan ini, bahasa pemrograman utama yang akan digunakan adalah Python untuk mengimplementasikan model GPT dan mengintegrasikannya dengan LangChain.

2.4 Pengujian

Pengujian aplikasi web question answering tentang peraturan perundang-undangan bidang pendidikan ini menggunakan BERT Score untuk memahami konteks dan makna pertanyaan secara mendalam, serta ROUGE Score untuk mengukur kesamaan dan kualitas jawaban yang dihasilkan, dengan tujuan memastikan bahwa jawaban yang diberikan akurat, relevan, dan sesuai dengan teks referensi yang diharapkan.

2.5 BERT Score

BERTScore adalah metrik evaluasi otomatis untuk generasi teks yang menggunakan embedding kontekstual dari model BERT (Bidirectional Encoder Representations from Transformers) [21]. BERTScore menghitung kesamaan antara dua kalimat berdasarkan kesamaan kosinus antara embedding token dalam kalimat-kalimat tersebut [22]. Ini dirancang untuk mengatasi keterbatasan dari metrik berbasis n-gram seperti BLEU yang sering gagal dalam menangkap sinonim dan urutan kata yang berbeda tetapi bermakna sama [23]. Dalam penghitungan skor lengkap, setiap token dalam x dicocokkan dengan token dalam $\hat{x}$ untuk menghitung recall, dan setiap token dalam $\hat{x}$ dicocokkan dengan token dalam x untuk menghitung precision. Pencocokan dilakukan secara greedy untuk memaksimalkan skor kesamaan pencocokan, di mana setiap token dicocokkan dengan token yang paling mirip dalam kalimat lainnya. Precision dan recall kemudian digabungkan untuk menghitung nilai F1 measure [21].

a. Menghitung Precision dan Recall:

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 9, No. 9, Month 2022, Page 999–999
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v9i9.9999

$$P_{Bert} = \frac{1}{|\hat{x}|}\sum_{\hat{x}_j \in \hat{x}} \underset{x_i \in x}{max} x_i^{\top} x_j \quad (1)$$

Rumus dalam menghitung precision dalam konteks ini mengukur seberapa akurat token-token dalam kalimat kandidat $\hat{x}$ mencocokkan token-token yang paling relevan dalam kalimat referensi x. Setiap token dalam kalimat kandidat akan dicocokkan dengan token dalam kalimat referensi berdasarkan kesamaan kosinus. Nilai Precision dihitung sebagai rata-rata dari nilai kesamaan kosinus tertinggi untuk semua token dalam kandidat $\hat{x}$. Semakin tinggi nilai Precision, semakin akurat kalimat kandidat dalam mencerminkan elemen-elemen penting dari kalimat referensi.

$$R_{Bert} = \frac{1}{|\hat{x}|}\sum_{\hat{x}_j \in \hat{x}} max\, x_i^{\top} x_j \quad (2)$$

Rumusan recall berfokus pada sejauh mana semua token dalam kalimat referensi x dapat dicocokkan dengan token dalam kalimat kandidat $\hat{x}$. Sama seperti precision, nilai recall dihitung berdasarkan nilai kesamaan kosinus tertinggi untuk setiap token dalam referensi, namun token dalam referensi yang menjadi acuan pencocokan. Semakin tinggi nilai Recall, semakin lengkap representasi kandidat terhadap seluruh informasi yang terdapat dalam referensi.

b. Menghitung Skor F1:

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

Rumusa f1 Score adalah metrik yang menggabungkan nilai Precision dan Recall menjadi satu nilai tunggal menggunakan rata-rata harmonis. Rumus ini dirancang untuk memberikan penilaian yang seimbang, terutama ketika terdapat perbedaan signifikan antara Precision dan Recall. Jika salah satu dari Precision atau Recall memiliki nilai rendah, maka F1 Score akan mencerminkan performa tersebut. Dengan demikian, F1 Score memberikan gambaran yang lebih holistik tentang seberapa baik model menjawab pertanyaan dengan mempertimbangkan akurasi dan kelengkapan jawaban.

2.6 ROUGEScore

ROUGE (Recall-Oriented Understudy for Gisting Evaluation) adalah sebuah paket evaluasi otomatis yang digunakan untuk menentukan kualitas sebuah ringkasan dengan membandingkannya dengan ringkasan lain yang dianggap ideal atau dibuat oleh manusia [24]. Metode ini menghitung jumlah unit yang tumpang tindih seperti n-gram, urutan kata, dan pasangan kata antara ringkasan yang dihasilkan oleh komputer dan ringkasan ideal yang dibuat oleh manusia. Dalam konteks ROUGE, Precision, Recall, dan F1 Score dihitung untuk mengukur kesamaan antara ringkasan yang dihasilkan oleh sistem (ringkasan kandidat) dan ringkasan referensi yang ideal (dibuat oleh manusia) [25]. Berikut penjelasan mengenai masing-masing metrik:

1. Precision: Mengukur proporsi n-gram dalam ringkasan kandidat yang juga muncul dalam ringkasan referensi.

$$Precision = \frac{Jumlah\ n{-}gram\ yang\ cocok}{Jumlah\ total\ n{-}gram\ dalam\ ringkasan\ kandidat} \quad (4)$$

Rumus precision dalam rouge score digunakan untuk mengukur seberapa besar proporsi n-gram (kelompok kata berurutan) dalam ringkasan kandidat yang juga terdapat dalam ringkasan referensi. Rumus ini menghitung persentase n-gram yang relevan dan sesuai dalam ringkasan kandidat dibandingkan dengan seluruh n-gram yang dihasilkan oleh sistem. Precision menjadi penting dalam menilai akurasi sistem, terutama ketika tujuan utamanya adalah menghasilkan ringkasan yang fokus dan tidak terlalu panjang. Semakin tinggi nilai Precision, semakin relevan informasi dalam ringkasan kandidat terhadap referensi.

2. Recall: Mengukur proporsi n-gram dalam ringkasan referensi yang juga muncul dalam ringkasan kandidat.

$$Recall = \frac{Jumlah\ n{-}gram\ yang\ cocok}{Jumlah\ total\ n{-}gram\ dalam\ ringkasan\ referensi} \quad (5)$$

Rumus recall bertujuan untuk mengukur sejauh mana ringkasan kandidat mencakup semua informasi penting dari ringkasan referensi. Rumus ini menghitung proporsi n-gram dalam referensi yang berhasil ditemukan dalam kandidat. Recall menjadi metrik yang sangat berguna ketika penekanan ada pada cakupan informasi, memastikan bahwa sistem tidak melewatkan bagian penting dari referensi. Semakin tinggi nilai Recall, semakin lengkap informasi dalam ringkasan kandidat dibandingkan dengan referensi.

3. F1 Score: Merupakan rata-rata harmonis dari Precision dan Recall, memberikan keseimbangan antara keduanya.

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (6)$$

Perbedaan antara ROUGE-1, ROUGE-2, dan ROUGE-L terletak pada n-gram yang mereka pertimbangkan dan cara mengukur kesamaan antara teks yang dihasilkan model dengan referensi manusia [24]. ROUGE-1 mempertimbangkan unigram (1-gram), yang berarti ia hanya memeriksa kesamaan kata per kata antara ringkasan kandidat dan referensi. Ini memungkinkan untuk mengukur kesamaan secara detail pada tingkat

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 9, No. 9, Month 2022, Page 999–999
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v9i9.9999

kata tunggal. ROUGE-2 mempertimbangkan bigram (2-gram), sehingga memeriksa kesamaan pasangan kata secara berdampingan. Ini memberikan pandangan yang lebih kontekstual, karena mempertimbangkan hubungan antara dua kata berurutan [24].

Sementara itu, ROUGE-L (Longest Common Subsequence) mencari urutan kata terpanjang yang sama antara ringkasan kandidat dan referensi. Ini tidak terbatas pada n-gram tertentu, melainkan mencari urutan kata yang paling panjang yang ada di kedua teks. Ini memberikan pemahaman yang lebih besar tentang kesamaan keseluruhan antara teks yang dihasilkan model dan referensi manusia. Jadi, meskipun ROUGE-1 fokus pada kata tunggal, ROUGE-2 mempertimbangkan pasangan kata, dan ROUGE-L memeriksa urutan kata terpanjang yang sama. Setiap metrik ROUGE memiliki sudut pandangnya sendiri dalam mengukur kualitas ringkasan terhadap referensi manusia [26].

2.5 Kesimpulan

Pada tahap kesimpulan, penelitian ini mencapai akhirnya dengan merangkum dan menarik simpulan dari seluruh hasil penelitian yang telah dilakukan

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Dokumen legal yang akan menjadi knowledge base dari sistem *questions answering* berbasis web ini, didapat melalui situs Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi (kemdikbudristek) yang dapat diakses melalui tautan berikut : https://jdih.kemdikbud.go.id/. Adapun produk hukum atau dokumen legal yang digunakan pada penelitian ini dapat dilihat pada tabel 1, sebagai berikut :

Tabel 1. Dokumen Legal Pendidikan

Jenis Dokumen	Nomor	Tahun	Tentang
Peraturan Perundang-undangan	20	2003	Sistem Pendidikan Nasional
Peraturan Perundang-undangan	14	2005	Guru dan Dosen
Peraturan Perundang-undangan	12	2012	Pendidikan Tinggi
Peraturan Perundang-undangan	11	2019	Sistem Nasional Ilmu Pengetahuan Dan Teknologi
Peraturan Perundang-undangan	23	2022	Pendidikan Dan Layanan Psikologi

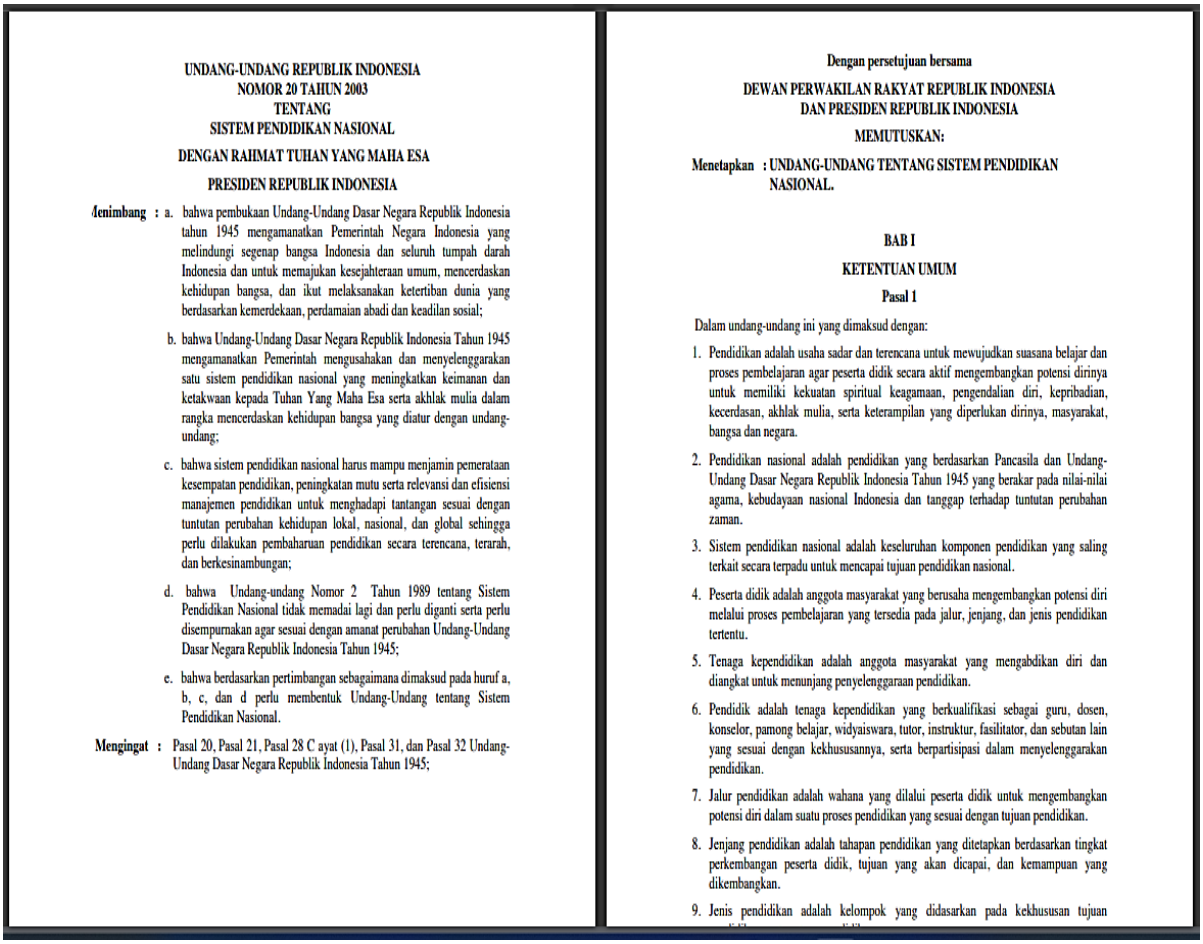
Dokumen legal yang tercantum pada Tabel 1, yaitu Undang-Undang Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional, menjadi salah satu data penting dalam penelitian ini. Hal ini disebabkan oleh perannya sebagai landasan hukum dan payung regulasi utama yang mengatur sistem pendidikan di Indonesia [27][28]. Undang-Undang Nomor 14 Tahun 2005 tentang Guru dan Dosen dipandang penting dalam penelitian ini karena mengatur secara tegas kualifikasi akademik, kompetensi, serta hak dan kewajiban dosen sebagai tenaga pendidik yang profesional.

Dosen tidak hanya dituntut untuk memiliki kualifikasi pendidikan tertentu, tetapi juga diharapkan mampu menjalankan peran mereka sebagai pendidik, pengajar, dan pembimbing bagi mahasiswa untuk menghasilkan lulusan berkualitas tinggi. Selain itu, UU ini juga menekankan kesejahteraan dosen, termasuk pemberian tunjangan profesi bagi yang telah bersertifikasi, yang bertujuan agar dosen dapat lebih fokus pada tugas utama mereka dalam mendidik mahasiswa tanpa perlu memikirkan tambahan pendapatan dari luar [27]. Sedangkan Undang-Undang Nomor 12 Tahun 2012, Undang-Undang Nomor 11 Tahun 2019, dan Undang-Undang Nomor 23 Tahun 2022 dipilih karena memiliki relevansi yang kuat dengan topik penelitian terkait pendidikan.

3.2 Analisa Dan Perancangan

a. Analisa Data
   Dokumen legal di bidang pendidikan yang telah dikumpulkan perlu dianalisis terlebih dahulu. Dari gambar 2 dibawah ini dokumen legal yang diunduh melalui situs Kemendikbud berformat PDF. Masalah muncul karena format ini seringkali kurang terstruktur dan menyulitkan pemahaman oleh sistem tanya jawab. Oleh karena itu, diperlukan konversi dokumen PDF ke format teks agar ukuran file lebih kecil, efisien, dan rapi, sehingga lebih mudah diolah oleh sistem tanya jawab (QAS). Hasil dari pemformatan dari pdf ke txt dapat dilihat pada gambar 3.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 9, No. 9, Month 2022, Page 999–999
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v9i9.9999



UNDANG-UNDANG REPUBLIK INDONESIA
NOMOR 20 TAHUN 2003
TENTANG
SISTEM PENDIDIKAN NASIONAL

DENGAN RAHMAT TUHAN YANG MAHA ESA

PRESIDEN REPUBLIK INDONESIA

Menimbang : a. bahwa pembukaan Undang-Undang Dasar Negara Republik Indonesia tahun 1945 mengamanatkan Pemerintah Negara Indonesia yang melindungi segenap bangsa Indonesia dan seluruh tumpah darah Indonesia dan untuk memajukan kesejahteraan umum, mencerdaskan kehidupan bangsa, dan ikut melaksanakan ketertiban dunia yang berdasarkan kemerdekaan, perdamaian abadi dan keadilan sosial;

b. bahwa Undang-Undang Dasar Negara Republik Indonesia Tahun 1945 mengamanatkan Pemerintah mengusahakan dan menyelenggarakan satu sistem pendidikan nasional yang meningkatkan keimanan dan ketakwaan kepada Tuhan Yang Maha Esa serta akhlak mulia dalam rangka mencerdaskan kehidupan bangsa yang diatur dengan undang-undang;

c. bahwa sistem pendidikan nasional harus mampu menjamin pemerataan kesempatan pendidikan, peningkatan mutu serta relevansi dan efisiensi manajemen pendidikan untuk menghadapi tantangan sesuai dengan tuntutan perubahan kehidupan lokal, nasional, dan global sehingga perlu dilakukan pembaharuan pendidikan secara terencana, terarah, dan berkesinambungan;

d. bahwa Undang-undang Nomor 2 Tahun 1989 tentang Sistem Pendidikan Nasional tidak memadai lagi dan perlu diganti serta perlu disempurnakan agar sesuai dengan amanat perubahan Undang-Undang Dasar Negara Republik Indonesia Tahun 1945;

e. bahwa berdasarkan pertimbangan sebagaimana dimaksud pada huruf a, b, c, dan d perlu membentuk Undang-Undang tentang Sistem Pendidikan Nasional.

Mengingat : Pasal 20, Pasal 21, Pasal 28 C ayat (1), Pasal 31, dan Pasal 32 Undang-Undang Dasar Negara Republik Indonesia Tahun 1945;

Dengan persetujuan bersama

DEWAN PERWAKILAN RAKYAT REPUBLIK INDONESIA
DAN PRESIDEN REPUBLIK INDONESIA

MEMUTUSKAN:

Menetapkan : UNDANG-UNDANG TENTANG SISTEM PENDIDIKAN NASIONAL.

BAB I

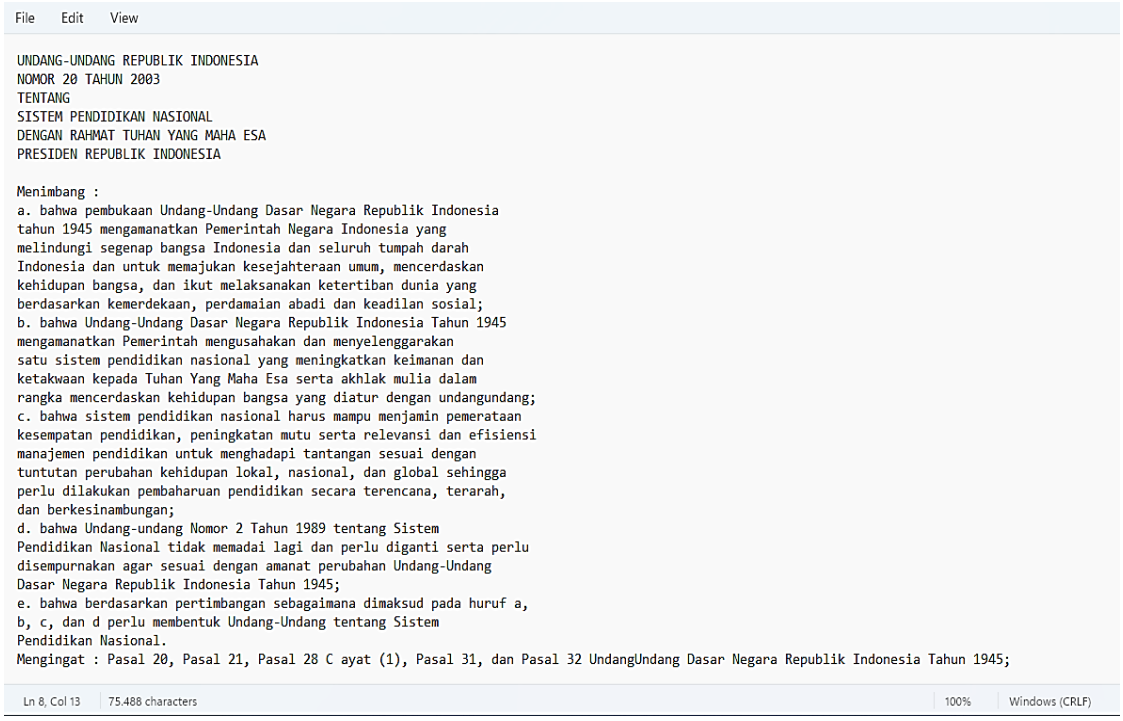
KETENTUAN UMUM

Pasal 1

Dalam undang-undang ini yang dimaksud dengan:

1. Pendidikan adalah usaha sadar dan terencana untuk mewujudkan suasana belajar dan proses pembelajaran agar peserta didik secara aktif mengembangkan potensi dirinya untuk memiliki kekuatan spiritual keagamaan, pengendalian diri, kepribadian, kecerdasan, akhlak mulia, serta keterampilan yang diperlukan dirinya, masyarakat, bangsa dan negara.
2. Pendidikan nasional adalah pendidikan yang berdasarkan Pancasila dan Undang-Undang Dasar Negara Republik Indonesia Tahun 1945 yang berakar pada nilai-nilai agama, kebudayaan nasional Indonesia dan tanggap terhadap tuntutan perubahan zaman.
3. Sistem pendidikan nasional adalah keseluruhan komponen pendidikan yang saling terkait secara terpadu untuk mencapai tujuan pendidikan nasional.
4. Peserta didik adalah anggota masyarakat yang berusaha mengembangkan potensi diri melalui proses pembelajaran yang tersedia pada jalur, jenjang, dan jenis pendidikan tertentu.
5. Tenaga kependidikan adalah anggota masyarakat yang mengabdikan diri dan diangkat untuk menunjang penyelenggaraan pendidikan.
6. Pendidik adalah tenaga kependidikan yang berkualifikasi sebagai guru, dosen, konselor, pamong belajar, widyaiswara, tutor, instruktur, fasilitator, dan sebutan lain yang sesuai dengan kekhususannya, serta berpartisipasi dalam menyelenggarakan pendidikan.
7. Jalur pendidikan adalah wahana yang dilalui peserta didik untuk mengembangkan potensi diri dalam suatu proses pendidikan yang sesuai dengan tujuan pendidikan.
8. Jenjang pendidikan adalah tahapan pendidikan yang ditetapkan berdasarkan tingkat perkembangan peserta didik, tujuan yang akan dicapai, dan kemampuan yang dikembangkan.
9. Jenis pendidikan adalah kelompok yang didasarkan pada kekhususan tujuan

Gambar 2. Dokumen Format PDF



File Edit View

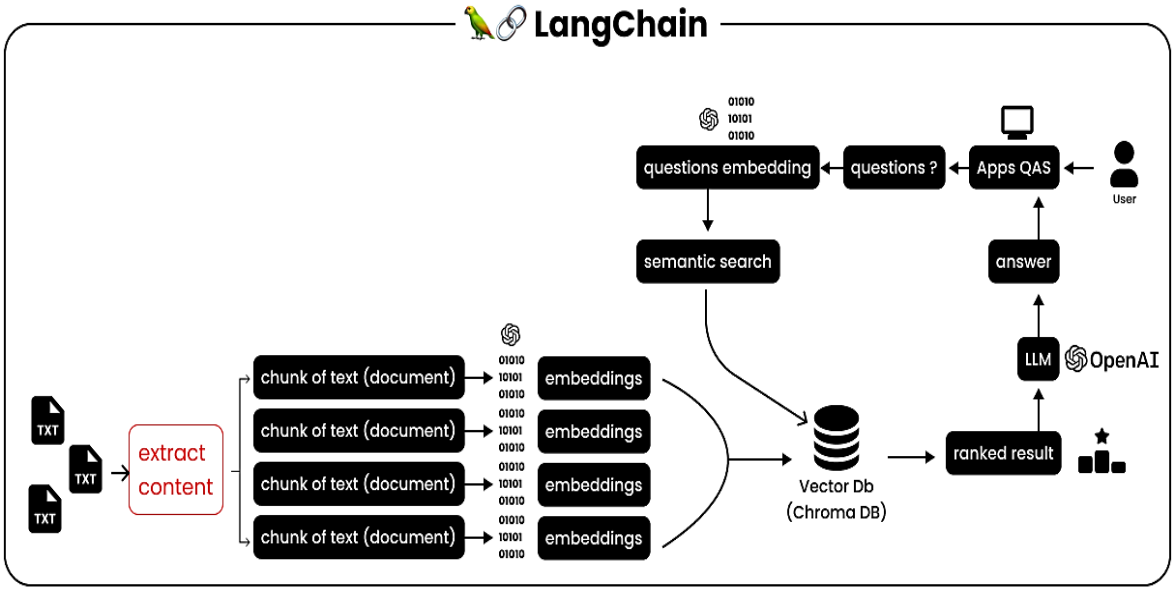
```
UNDANG-UNDANG REPUBLIK INDONESIA
NOMOR 20 TAHUN 2003
TENTANG
SISTEM PENDIDIKAN NASIONAL
DENGAN RAHMAT TUHAN YANG MAHA ESA
PRESIDEN REPUBLIK INDONESIA

Menimbang :
a. bahwa pembukaan Undang-Undang Dasar Negara Republik Indonesia
tahun 1945 mengamanatkan Pemerintah Negara Indonesia yang
melindungi segenap bangsa Indonesia dan seluruh tumpah darah
Indonesia dan untuk memajukan kesejahteraan umum, mencerdaskan
kehidupan bangsa, dan ikut melaksanakan ketertiban dunia yang
berdasarkan kemerdekaan, perdamaian abadi dan keadilan sosial;
b. bahwa Undang-Undang Dasar Negara Republik Indonesia Tahun 1945
mengamanatkan Pemerintah mengusahakan dan menyelenggarakan
satu sistem pendidikan nasional yang meningkatkan keimanan dan
ketakwaan kepada Tuhan Yang Maha Esa serta akhlak mulia dalam
rangka mencerdaskan kehidupan bangsa yang diatur dengan undangundang;
c. bahwa sistem pendidikan nasional harus mampu menjamin pemerataan
kesempatan pendidikan, peningkatan mutu serta relevansi dan efisiensi
manajemen pendidikan untuk menghadapi tantangan sesuai dengan
tuntutan perubahan kehidupan lokal, nasional, dan global sehingga
perlu dilakukan pembaharuan pendidikan secara terencana, terarah,
dan berkesinambungan;
d. bahwa Undang-undang Nomor 2 Tahun 1989 tentang Sistem
Pendidikan Nasional tidak memadai lagi dan perlu diganti serta perlu
disempurnakan agar sesuai dengan amanat perubahan Undang-Undang
Dasar Negara Republik Indonesia Tahun 1945;
e. bahwa berdasarkan pertimbangan sebagaimana dimaksud pada huruf a,
b, c, dan d perlu membentuk Undang-Undang tentang Sistem
Pendidikan Nasional.
Mengingat : Pasal 20, Pasal 21, Pasal 28 C ayat (1), Pasal 31, dan Pasal 32 UndangUndang Dasar Negara Republik Indonesia Tahun 1945;
```

Ln 8, Col 13 | 75.488 characters | 100% | Windows (CRLF)

Gambar 3. Dokumen Format TXT

b. Perancangan *Questions Answering*
Tahapan perancangan sistem *questions answering* (qas) berbasis web menggunakan kerangka kerja LangChain dapat dilihat pada Gambar 4 dibawah.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 9, No. 9, Month 2022, Page 999–999
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v9i9.9999



Gambar 4. Tahapan Perancangan Sistem QAS

Kerangka kerja LangChain menyediakan langkah-langkah sistematis dalam merancang aplikasi dengan memanfaatkan model bahasa besar (LLM) [3]. Berikut adalah penjelasan setiap langkah dalam merancang sistem penjawaban pertanyaan (QAS) pada penelitian ini:

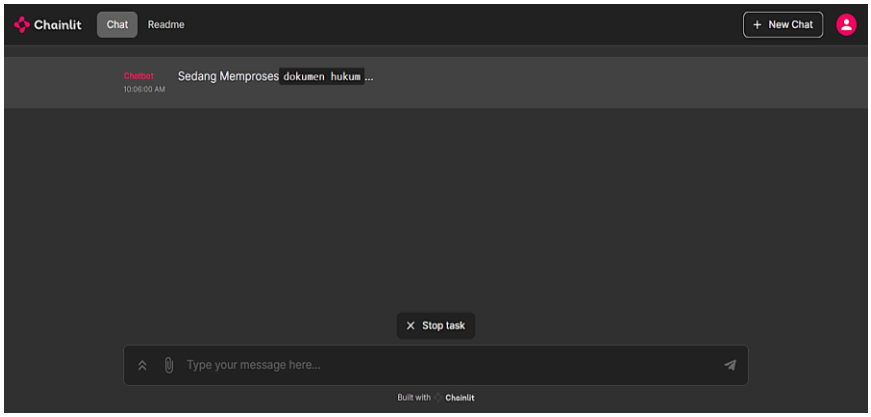
1. Tahapan pertama adalah memuat dokumen legal bidang pendidikan ke dalam kerangka kerja langchain, nantinya dokumen ini akan di chunk atau dipotong menjadi beberapa bagian.
2. Selanjutnya tahapan kedua adalah pemotongan teks, proses pemotongan dokumen dengan menggunakan LangChain Framework meningkatkan efisiensi pengambilan informasi dengan membagi dokumen menjadi potongan-potongan berdasarkan karakteristik teks. Proses ini memfokuskan pada segmen yang relevan dan menjaga koherensi konteks di antara potongan-potongan tersebut, sehingga memungkinkan pengambilan informasi yang lebih tepat dan efektif [5]. Potongan-potongan dibuat dengan membagi teks berdasarkan pemisah yang ditentukan, seperti tanda titik ('.'), dengan ukuran potongan 1000 karakter. Jika potongan melebihi ukuran yang ditentukan, potongan tersebut tetap dibiarkan apa adanya, sementara potongan-potongan yang lebih kecil dapat digabungkan dengan yang berdekatan. Overlap 100 karakter memastikan aliran informasi yang mulus di antara potongan-potongan yang berdekatan.
3. Pada tahap ketiga, sekarang kita perlu memasukkan potongan-potongan ini ke dalam indeks sehingga kita dapat mengambilnya dengan mudah ketika kita ingin menjawab pertanyaan pada dokumen ini. Kami menggunakan embeddings dan penyimpanan vektor untuk tujuan ini. Untuk penelitian kami, kami menggunakan ChromaDB, sebuah basis data embedding open-source. ChromaDB mendukung penyimpanan lokal dan menyediakan wrapper yang mudah digunakan untuk model embedding populer seperti Sentence Transformers, API embedding OpenAI, dan instructor-embeddings. Selain itu, ChromaDB juga memungkinkan penggunaan model embedding kustom [5].
4. Setelah tahap embedding sudah dilakukan sekarang vector db sudah berisi sumber konteks dokumen legal, selanjutnya user bisa mengajukan pertanyaan dari dokumen yang sudah dimuat sebelumnya pada tahap 1.
5. Pada tahap 5 pertanyaan yang diajukan akan melewati proses embedding oleh llm openai, yang nantinya pertanyaan tersebut akan dicocokan dengan sumber konteks
6. Tahapan nomor 6 akan di lakukan semantic search atau pencarian yang semantik dari pertanyaan yang di ajukan ke dalam vector db.
7. Selanjutnya tahapan terakhir yang ke-7, jika jawaban ditemukan akan dilakukan pemeringkatan terlebih dahulu untuk memastikan kesesuain jawaban dengan pertanyaan, setelah itu llm akan membuat jawaban untuk pertanyaan yang diajukan oleh user pada tahap ke-4.

3.3 Implementasi

Pada sub bab ini, akan dibahas mengenai implementasi aplikasi question answering. Implementasi ini mencakup berbagai tahapan yang ditunjukkan melalui screenshot aplikasi question answering, mulai dari tampilan awal aplikasi, tampilan prompt untuk pertanyaan, tampilan jawaban yang diberikan, hingga tampilan sumber informasi yang digunakan untuk menjawab pertanyaan. Mari kita lihat lebih detail setiap tahapannya.

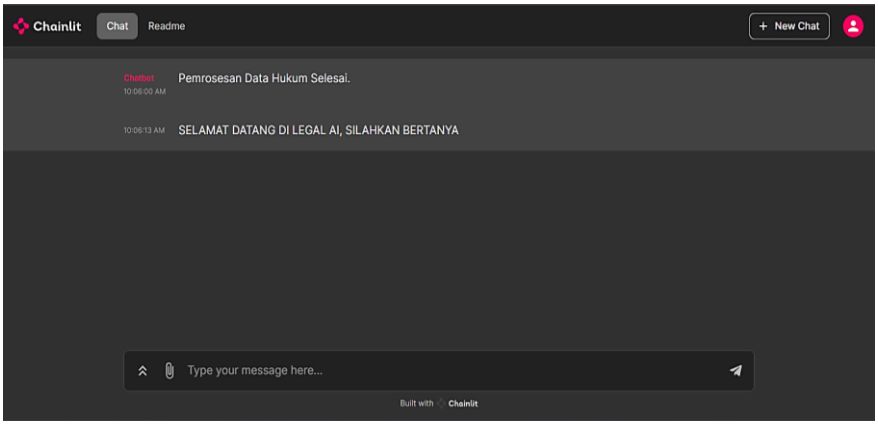
a. Tampilan awal question answering
   Gambar 5 menampilkan sistem diakses untuk pertama kali, aplikasi question answering (QAS) akan menjalankan proses pra-pengolahan dokumen hukum untuk mempersiapkan basis data yang diperlukan.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 9, No. 9, Month 2022, Page 999–999
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v9i9.9999



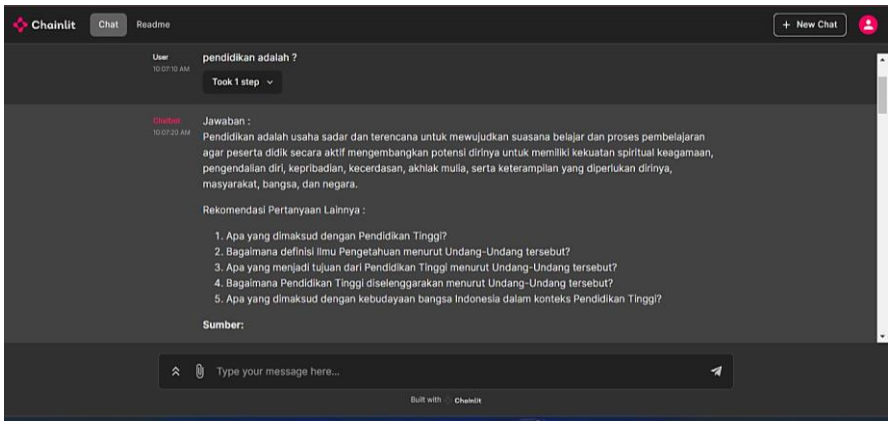
Gambar 5. Tampilan awal question answering

b. Tampilan prompt pertanyaan
Setelah proses pra-pengolahan dokumen hukum selesai, sebuah prompt akan muncul, memungkinkan pengguna untuk mengajukan pertanyaan yang relevan dengan dokumen hukum yang telah diproses dapat dilihat pada gambar 6.



Gambar 6. Tampilan prompt pertanyaan

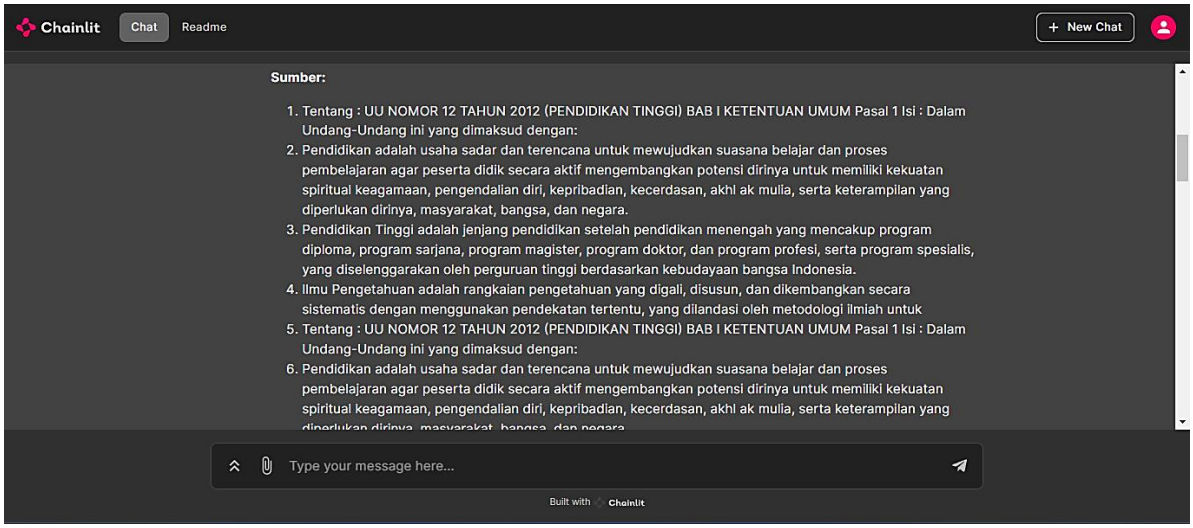
c. Tampilan Jawaban
Setelah pertanyaan diajukan, aplikasi akan menampilkan jawaban yang sesuai. Jawaban tersebut dilengkapi dengan rekomendasi pertanyaan terkait yang berhubungan dengan jawaban yang diberikan dan pertanyaan sebelumnya dapat dilihat pada gambar 7.



Gambar 7. Tampilan Jawaban

d. Tampilan Sumber Informasi
Pada gambar 8 untuk setiap jawaban, aplikasi question answering akan menampilkan sumber informasi yang digunakan. Informasi ini mencakup jenis undang-undang, bab, dan pasal yang relevan dengan jawaban tersebut.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 9, No. 9, Month 2022, Page 999–999
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v9i9.9999



Gambar 8. Tampilan Sumber Informasi

3.4 Pengujian Dan Evaluasi QAS

Pengujian dilakukan dengan menggunakan metode evaluasi BERTScore dan ROGUEScore, bertujuan untuk menentukan sejauh mana model dapat menjawab pertanyaan dengan akurat berdasarkan referensi yang diberikan. Pengujian dilakukan dengan menggunakan 10 pertanyaan yang telah disiapkan sebelumnya, beserta prediksi jawabannya dapat dilihat pada tabel 2 dibawah ini.

Tabel 2. Daftar Pertanyaan Pengujian

No	Pertanyaan
1.	Sistem pendidikan nasional adalah ?
2.	Apakah orang tua berhak berperan dalam pendidikan anak ?
3.	Guru besar atau profesor adalah ?
4.	Kedudukan guru dan dosen sebagai tenaga profesional bertujuan untuk ?
5.	Pendidikan Tinggi adalah ?
6.	Pendidikan Tinggi diselenggarakan dengan prinsip ?
7.	Ilmu Pengetahuan dan Teknologi berperan sebagai ?
8.	Ilmu Pengetahuan dan Teknologi dikembangkan melalui ?
9.	Pendidikan Psikologi adalah ?
10.	STR sebagaimana dimaksud dalam Pasal 15 tidak berlaku apabila ?

Tabel 2 menampilkan 10 pertanyaan yang dirancang secara khusu berdasarkan dokumen legal di bidang pendidikan. Setiap dokumen legal menjadi sumber untuk 2 pertanyaan, sehingga total terdapat 10 pertanyaan. Pemilihan pertanyaan dilakukan dengan mempertimbangkan cakupan isi dan relevansi terhadap dokumen legal yang digunakan. Adapun referensi dan dan prediksi dari pertanyaan dapat dilihat pada tabel 3.

Tabel 3. Contoh Pertanyaan Dengan Referensi dan Prediksi

No	Pertanyaan	Referensi	Prediksi
1.	Sistem pendidikan nasional adalah ?	Sistem pendidikan nasional adalah keseluruhan komponen pendidikan yang saling terkait secara terpadu untuk mencapai tujuan pendidikan nasional.	Sistem pendidikan nasional adalah keseluruhan komponen pendidikan yang saling terkait secara terpadu untuk mencapai tujuan pendidikan nasional.
2.	Apakah orang tua berhak berperan dalam pendidikan anak ?	Orang tua berhak berperan serta dalam memilih satuan pendidikan dan memperoleh informasi tentang perkembangan pendidikan anaknya. Orang tua dari anak usia wajib belajar, berkewajiban memberikan pendidikan dasar kepada anaknya.	Ya, orang tua memiliki peran dalam pendidikan anak sesuai dengan UU Nomor 20 Tahun 2003. Mereka berhak berperan serta dalam memilih satuan pendidikan untuk anak dan memperoleh informasi tentang perkembangan pendidikan anaknya. Selain itu, orang tua dari anak usia wajib belajar juga berkewajiban memberikan pendidikan dasar kepada anaknya.
3.	Guru besar atau profesor adalah ?	Guru besar atau profesor yang selanjutnya disebut profesor adalah jabatan fungsional tertinggi bagi dosen	Profesor adalah jabatan fungsional tertinggi bagi dosen yang masih mengajar di lingkungan satuan pendidikan tinggi.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 9, No. 9, Month 2022, Page 999–999
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v9i9.9999

No	Pertanyaan	Referensi	Prediksi
		yang masih mengajar di lingkungan satuan pendidikan tinggi. Guru besar adalah jabatan akademik tertinggi pada satuan pendidikan tinggi yang memiliki kewenangan membimbing calon doktor	Sementara itu, guru besar adalah jabatan akademik tertinggi pada satuan pendidikan tinggi yang memiliki kewenangan membimbing calon doktor.

Tabel 3 menyajikan contoh pertanyaan yang diujikan bersama dengan referensi jawaban dan prediksi jawaban yang dihasilkan oleh sistem. Referensi merupakan jawaban asli yang diambil langsung dari sumber konteks, yaitu dokumen legal atau undang-undang yang digunakan sebagai dasar pengujian. Prediksi adalah jawaban yang dihasilkan oleh sistem berbasis question answering sebagai respons terhadap pertanyaan yang diajukan. Pada tabel 4 menampilkan hasil evaluasi performa model menggunakan metrik BERTScore dan ROUGE. Hasil perhitungan ini mencakup nilai precision, recall, dan F1 score untuk setiap pertanyaan yang diuji.

Tabel 4. Hasil Perhitungan Dan Evalusi BERTScore Dan ROUGEScore

	BERTScore (%)			ROUGE-1 (%)			ROUGE-2 (%)			ROUGE-L (%)		
Pertanyaan	P	R	F1	P	R	F1	P	R	F1	P	R	F1
1.	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %
2.	86 %	93 %	90 %	60 %	100 %	75 %	51 %	86 %	64 %	60 %	100 %	75 %
3.	92 %	91 %	91 %	93 %	81 %	87 %	90 %	78 %	84 %	93 %	81 %	87 %
4.	98 %	98 %	98 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %
5.	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %
6.	97 %	97 %	97 %	93 %	93 %	93 %	89 %	89 %	89 %	93 %	93 %	93 %
7.	81 %	87 %	84 %	45 %	86 %	59 %	40 %	77 %	52 %	45 %	86 %	60 %
8.	75 %	78 %	77 %	44 %	36 %	40 %	23 %	19 %	21 %	38 %	32 %	35 %
9.	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %
10.	80 %	84 %	82 %	50 %	75 %	60 %	38 %	58 %	46 %	50 %	75 %	60 %
Rata - Rata	80 %	94 %	93%	78 %	89 %	79 %	69 %	68 %	67 %	69 %	83 %	73 %

Keterangan :

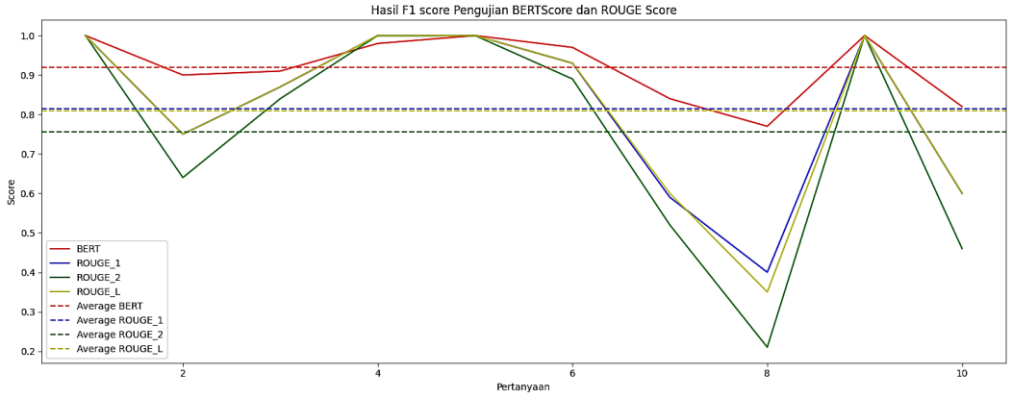
P : Precision (Presisi) - Mengukur ketepatan model dalam memberikan jawaban yang benar dibandingkan dengan semua jawaban yang diberikan.

R : Recall - Mengukur kemampuan model dalam menemukan semua jawaban yang benar dari semua jawaban yang mungkin ada.

F1 : F1 Score - Menggabungkan Precision dan Recall dalam satu metrik yang dihitung sebagai rata-rata harmonis dari keduanya

Berdasarkan hasil evaluasi pada tabel 4, metode BERTScore menunjukkan hasil yang lebih unggul dibandingkan ROUGE Score dalam menilai kesesuaian antara referensi dan prediksi dari sistem. Beberapa pertanyaan mendapatkan skor 100%, yang berarti bahwa referensi dan prediksi dari sistem sepenuhnya sesuai dan layak. ROUGE Score, yang mengandalkan kesamaan leksikal, cenderung mengukur kesamaan berdasarkan token atau urutan kata yang tepat, yang dapat mengabaikan kesamaan semantik yang ada ketika kata-kata berbeda digunakan untuk menyampaikan arti yang sama.

Sebaliknya, BERTScore menggunakan model transformator seperti BERT untuk menghasilkan embedding kontekstual dari token, yang memungkinkan pengukuran kesamaan berdasarkan konteks semantik keseluruhan dari teks tersebut. Hal ini berarti BERTScore lebih mampu menangkap nuansa makna yang lebih dalam dan variabilitas dalam bahasa alami, membuatnya lebih sesuai untuk evaluasi model tanya jawab dan tugas pemahaman bahasa alami lainnya yang kompleks.



Gambar 9. Grafik Hasil F1 Score Metode BERTScore dan ROUGEScore

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 9, No. 9, Month 2022, Page 999−999
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v9i9.9999

Grafik pada gambar 9 memperlihatkan perbandingan F1 score dari metode BERTScore dan ROUGE Score (ROUGE-1, ROUGE-2, dan ROUGE-L) pada serangkaian pengujian. BERT secara konsisten mencetak skor lebih tinggi dibandingkan dengan ROUGE pada hampir semua pertanyaan, kecuali dalam beberapa kasus di mana skor ROUGE mendekati skor BERT, terutama pada ROUGE-1 dan ROUGE-L. Rata-rata F1 score untuk BERT (0.929) jauh lebih tinggi dibandingkan dengan rata-rata untuk ROUGE-1 (0.814), ROUGE-2 (0.756), dan ROUGE-L (0.810).

ROUGE-2 dan ROUGE-L menunjukkan variasi yang lebih signifikan di seluruh pertanyaan, terutama untuk skor yang lebih rendah. Penurunan yang terlihat pada ROUGE-2 (misalnya, pada pertanyaan ke-8) dan ROUGE-L (misalnya, pada pertanyaan ke-8 dan ke-10) menunjukkan bahwa metrik ini mungkin lebih sensitif atau kurang stabil untuk jenis konten tertentu.

Skor BERT relatif konsisten, dengan sebagian besar skor mendekati nilai maksimum (1.00) atau sedikit di bawahnya, yang menunjukkan bahwa BERTScore mungkin lebih andal dalam menilai kualitas teks dalam berbagai kondisi pengujian.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan aplikasi web question answering yang memanfaatkan framework LangChain dan model LLM OpenAI untuk menjawab pertanyaan berdasarkan dokumen legal di bidang pendidikan. Hasil penelitian menunjukkan bahwa metode pengujian BERTScore lebih unggul dibandingkan ROUGE Score dalam mengukur kesesuaian antara jawaban sistem dan referensi, karena BERTScore mampu menangkap nuansa makna yang lebih dalam dan variabilitas bahasa alami lebih baik daripada ROUGE Score yang lebih bergantung pada kesamaan leksikal. Namun, penelitian ini menghadapi beberapa keterbatasan, seperti kebutuhan akan upaya manual dalam memastikan integritas dan keterbacaan data selama konversi dokumen legal dari format PDF ke teks, serta penggunaan sumber daya komputasi yang besar untuk implementasi LangChain dalam pemotongan dan pengolahan teks. Meskipun BERTScore memberikan hasil yang baik, beberapa kasus menunjukkan kesamaan semantik yang tidak sepenuhnya tertangkap oleh model, mengindikasikan perlunya model yang lebih canggih atau penyesuaian dalam preprocessing data. Penelitian lanjutan dapat memperbaiki aspek ini dengan menerapkan teknik konversi PDF yang lebih otomatis, mengeksplorasi model bahasa alami yang lebih kompleks, serta menguji sistem pada dataset yang lebih besar dan beragam untuk meningkatkan generalisasi. Secara keseluruhan, aplikasi yang dikembangkan menunjukkan potensi signifikan dalam memfasilitasi akses dan pemahaman dokumen legal pendidikan, dengan prospek menjadi alat yang berguna bagi praktisi pendidikan dan hukum melalui pengembangan lebih lanjut.

REFERENCES

[1] L. Beurer-Kellner, M. Fischer, and M. Vechev, “Prompting Is Programming: A Query Language for Large Language Models,” *Proc. ACM Program. Lang.*, vol. 7, no. June, pp. 186:2-186:3, 2023, doi: 10.1145/3591300.

[2] S. Ott *et al.*, “ThoughtSource: A central hub for large language model reasoning data,” *Sci. Data*, vol. 10, no. 1, pp. 1–13, 2023, doi: 10.1038/s41597-023-02433-3.

[3] 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, 2023, doi: 10.59287/icaens.1127.

[4] K. Pandya and B. V. Mahavidyalaya, “Automating Customer Service using LangChain,” *Comput. Lang. (cs.CL); Comput. Soc. (cs.CY); Mach. Learn.*, vol. 1, pp. 28–31, 2023, doi: https://doi.org/10.48550/arXiv.2310.05421.

[5] S. K. Nigam, S. K. Mishra, A. K. Mishra, N. Shallum, and A. Bhattacharya, “Legal Question-Answering in the Indian Context: Efficacy, Challenges, and Potential of Modern AI Models,” *Comput. Lang. (cs.CL); Artif. Intell.*, vol. 1–2, pp. 1–15, 2023, doi: https://doi.org/10.48550/arXiv.2309.14735.

[6] R. Nakano *et al.*, “WebGPT: Browser-assisted question-answering with human feedback,” *Comput. Lang. (cs.CL); Artif. Intell. (cs.AI); Mach. Learn.*, vol. 1–3, pp. 2–32, 2021, doi: https://doi.org/10.48550/arXiv.2112.09332.

[7] T. Lubiana *et al.*, “Ten quick tips for harnessing the power of ChatGPT in computational biology,” *PLOS Comput. Biol.*, vol. 19, no. 8, p. e1011319, Aug. 2023, doi: 10.1371/journal.pcbi.1011319.

[8] 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.*, vol. 05, no. 06, pp. 3980–3983, 2023, doi: 10.56726/irjmets42630.

[9] L. Floridi and M. Chiriatti, “GPT-3: Its Nature, Scope, Limits, and Consequences,” *Minds Mach.*, vol. 30, no. 4, pp. 681–694, 2020, doi: 10.1007/s11023-020-09548-1.

[10] T. Kojima, S. S. Gu, M. Reid, Y. Matsuo, and Y. Iwasawa, “Large Language Models are Zero-Shot Reasoners,” *Adv. Neural Inf. Process. Syst.*, vol. 35, no. NeurIPS, 2022.

[11] H. K. Aroral, “Waterfall Process Operations in the Fast-paced World: Project Management Exploratory

ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 9, No. 9, Month 2022, Page 999–999
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v9i9.9999

Analysis," *Int. J. Appl. Bus. Manag. Stud.*, vol. 6, no. 1, p. 2021, 2021.
[12] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: state of the art, current trends and challenges," *Multimed. Tools Appl.*, vol. 82, no. 3, pp. 3713–3744, 2023, doi: 10.1007/s11042-022-13428-4.
[13] B. A. Andrei, A. C. Casu-Pop, S. C. Gheorghe, and C. A. Boiangiu, "a Study on Using Waterfall and Agile Methods in Software Project Management," *J. Inf. Syst. Oper. Manag.*, pp. 125–235, 2019.
[14] I. Tri Julianto, D. Kurniadi, Y. Septiana, and A. Sutedi, "Alternative Text Pre-Processing using Chat GPT Open AI," *J. Nas. Pendidik. Tek. Inform.*, vol. 12, no. 1, pp. 67–77, 2023, doi: 10.23887/janapati.v12i1.59746.
[15] G. Chowdhury, "Natural language processing . Annual Review of This is an author-produced version of a paper published in The Annual Review of Information Science and Technology ISSN 0066-4200 . This version has been peer-reviewed , but does not," *Annu. Rev. Inf. Sci. Technol.*, vol. 37, pp. 51–89, 2003.
[16] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, vol. 1, no. Mlm, pp. 4171–4186, 2019.
[17] T. B. Brown *et al.*, "Language models are few-shot learners," *Adv. Neural Inf. Process. Syst.*, vol. 2020-Decem, 2020.
[18] Z. A. Yilmaz, S. Wang, W. Yang, H. Zhang, and J. Lin, "Birch: Applying BERT to document retrieval," *EMNLP-IJCNLP 2019 - 2019 Conf. Empir. Methods Nat. Lang. Process. 9th Int. Jt. Conf. Nat. Lang. Process. Proc. Syst. Demonstr.*, pp. 19–24, 2020.
[19] W. Sakata, R. Tanaka, T. Shibata, and S. Kurohashi, "FAQ retrieval using query-question similarity and BERT-based query-answer relevance," *SIGIR 2019 - Proc. 42nd Int. ACM SIGIR Conf. Res. Dev. Inf. Retr.*, pp. 1113–1116, 2019, doi: 10.1145/3331184.3331326.
[20] Y. Qiao, C. Xiong, Z. Liu, and Z. Liu, "Understanding the Behaviors of BERT in Ranking," 2019, [Online]. Available: http://arxiv.org/abs/1904.07531.
[21] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, and Y. Artzi, "Bertscore: Evaluating Text Generation With Bert," *8th Int. Conf. Learn. Represent. ICLR 2020*, pp. 1–43, 2020.
[22] J. Risch, T. Möller, J. Gutsch, and M. Pietsch, "Semantic Answer Similarity for Evaluating Question Answering Models," *Proc. 3rd Work. Mach. Read. Quest. Answering, MRQA 2021*, pp. 149–157, 2021, doi: 10.18653/v1/2021.mrqa-1.15.
[23] A. Chen, G. Stanovsky, S. Singh, and M. Gardner, "Evaluating question answering evaluation," *MRQA@EMNLP 2019 - Proc. 2nd Work. Mach. Read. Quest. Answering*, pp. 119–124, 2019, doi: 10.18653/v1/d19-5817.
[24] M. Barbella and G. Tortora, "Rouge Metric Evaluation for Text Summarization Techniques," *SSRN Electron. J.*, 2022, doi: 10.2139/ssrn.4120317.
[25] G. Tsuchiya, "Postmortem Angiographic Studies on the Intercoronary Arterial Anastomoses.: Report I. Studies on Intercoronary Arterial Anastomoses in Adult Human Hearts and the Influence on the Anastomoses of Strictures of the Coronary Arteries.," *Jpn. Circ. J.*, vol. 34, no. 12, pp. 1213–1220, 1971, doi: 10.1253/jcj.34.1213.
[26] W. Tay, A. Joshi, X. Zhang, S. Karimi, and S. Wan, "Red-faced ROUGE: Examining the Suitability of ROUGE for Opinion Summary Evaluation," *Proc. 17th Annu. Work. Australas. Lang. Technol. Assoc.*, pp. 52–60, 2019, [Online]. Available: https://www.aclweb.org/anthology/U19-1008.
[27] Muchammad Catur Rizky, Rohman Hakim, Miftakhul Anam, Moch Nur Alim, and Wahyu Suhartatik, "Implementasi Undang-Undang Nomor 14 Tahun 2005 Tentang Guru dan Dosen terhadap Kesejahteraan Dosen Profesional di Universitas Sunan Giri Surabaya," *J. Kolaboratif Sains*, vol. 5, no. 8, pp. 561–569, 2022, doi: 10.56338/jks.v5i8.2734.