

SISTEM TANYA JAWAB UNTUK PERTANYAAN APA MENGUNAKAN METODE PENDEKATAN JARAK

Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Jurusan Teknik Informatika



Oleh :

Zafira Galea
NIM : 09021281924049

**Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2023**

LEMBAR PENGESAHAN SKRIPSI

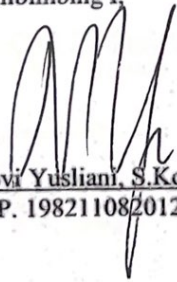
SISTEM TANYA JAWAB UNTUK PERTANYAAN APA
MENGUNAKAN METODE PENDEKATAN JARAK

Oleh :

Zafira Galea
NIM : 09021281924049

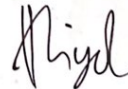
Palembang, 05 Juni 2023

Pembimbing I,



Novi Yusliani, S.Kom., M.T.
NIP. 198211082012122001

Pembimbing II,



Junia Kurniati, M.Kom.
NIK. 1671046606890018

Mengetahui,
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom.
NIP. 19781222200642003

TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI

Pada hari Senin tanggal 31 Juli 2023 telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya

Nama : Zafira Golea

NIM : 09021281924049

Judul : SISTEM TANYA JAWAB UNTUK PERTANYAAN APA
MENGGUNAKAN METODE PENDEKATAN JARAK

dan dinyatakan LULUS.

1. Ketua Penguji

Mastam Rizka Mariska, M.T.
NIP. 198603212019032001

2. Penguji

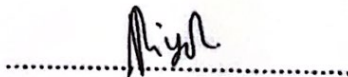
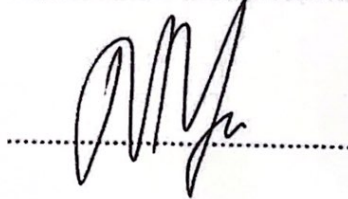
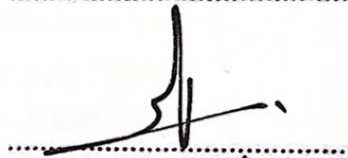
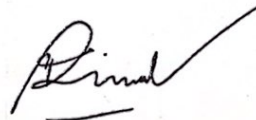
Dr. Abdiansah, S.Kom., M.Cs.
NIP. 198410012009121003

3. Pembimbing I

Novi Yudianti, S.Kom., M.T.
NIP. 198211082012122001

4. Pembimbing II

Junia Kurnjati, M.Kom.
NIK. 1671045606890018



Mengetahui,
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Zafira Galea
NIM : 09021281924049
Program Studi : Teknik Informatika
Judul Skripsi : Sistem Tanya Jawab Untuk Pertanyaan Apa
Menggunakan Metode Pendekatan Jarak

Hasil Pengecekan Software (iThenticate/Turnitin) : 12%

Menyatakan bahwa laporan proyek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan / plagiat. Apabila ditemukan unsur penjiplakan / plagiat dalam laporan proyek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Palembang, 31 Juli 2023



Zafira Galea

NIM. 09021281924049

MOTTO DAN PERSEMBAHAN

"Live with things that are right and just go with the flow."

Kupersembahkan karya tulis ini kepada :

- Keluarga Besarku
- Teman Seperjuangan
- Fakultas Ilmu Komputer
- Universitas Sriwijaya

ABSTRACT

Question Answering System (QAS) is a system designed to generate information quickly in the form of natural language. The information produced by this system consists of short text snippets as answers to user questions. In this research on Question Answering System (QAS), users can obtain answers related to non-factoid information questions. The generated information is close domain, sourced from a segment of biology learning information obtained from the best documents. The information retrieval process involves preprocessing to obtain keywords and important words, calculating TF-IDF, and using three distance-approach algorithms: Manhattan Distance, Euclidean Distance, and Cosine Similarity to retrieve the best documents and answers. In the testing phase, 110 question-answer pairs and 61 documents containing answers were used. The best accuracy was achieved by the Manhattan Distance algorithm, with 0.9% in test 1, where only the definition answers were considered. The best accuracy for Euclidean Distance was 36.36%, and for Cosine Similarity, it was 37.27% in test 4, where both the definition answers and keywords with a 40% threshold were considered, obtaining a reasonably good result. This research demonstrates the capability to perform question-answering for non-factoid questions, and it highlights how the preprocessing process, the type of distance-approach algorithms used, and the testing methods can affect the accuracy of the system.

Keywords: *QAS, Manhattan Distance, Euclidean Distance, Cosine Similarity.*

ABSTRAK

Sistem Tanya Jawab merupakan sistem yang dirancang untuk menghasilkan informasi secara cepat dalam bentuk *natural language* (bahasa alami). Informasi yang dihasilkan sistem ini berupa kutipan teks singkat sebagai jawaban untuk menjawab pertanyaan pengguna. Pada penelitian Sistem Tanya Jawab ini, pengguna akan mendapatkan jawaban seputaran informasi mengenai definisi *non-factoid* untuk pertanyaan apa. Informasi yang dihasilkan berupa *close domain*, informasi berasal dari sepenggal informasi pembelajaran biologi yang diperoleh dari dokumen terbaik. Pencarian informasi melalui proses pra-pengolahan dalam memperoleh kata kunci dan kata-kata penting, perhitungan TF-IDF, dan perhitungan 3 algoritma pendekatan jarak *Manhattan Distance*, *Euclidean Distance*, dan *Cosine Similarity* untuk memperoleh dokumen dan jawaban terbaik. Pada pengujian 110 pasang pertanyaan dan jawaban apa, serta 61 dokumen yang mengandung jawaban, didapatkan nilai akurasi terbaik dari algoritma *Manhattan Distance* sebesar 0.9% pada percobaan pengujian 1 dengan tanpa mengambil jawaban definisi saja. Sedangkan nilai akurasi terbaik pada *Euclidean Distance* 36.36%, dan *Cosine Similarity* 37.27% terdapat pada percobaan pengujian 4 dengan mengambil jawaban definisi dan mengambil kata kunci dengan threshold 40% yang memperoleh nilai cukup baik. Hasil penelitian ini dapat melakukan proses tanya jawab untuk pertanyaan apa *non-factoid* serta proses pra-pengolahan, jenis algoritma metode pendekatan jarak, dan metode pengujian akan mempengaruhi nilai akurasi.

Kata Kunci: QAS, *Manhattan Distance*, *Euclidean Distance*, *Cosine Similarity*

KATA PENGANTAR

Puji syukur atas kehadiran Allah SWT yang telah memberikan Rahmat, hidayah, dan karunia_Nya sehingga penulis dapat menyelesaikan Tugas Akhir ini dengan judul “Sistem Tanya Jawab Untuk Pertanyaan Apa Menggunakan Metode Pendekatan Jarak”.

Penulisan skripsi ini tidak akan selesai pada waktunya dan tidak akan terwujud tanpa dukungan, bimbingan, dan bantuan dari berbagai pihak. Oleh karena itu, ungkapan rasa syukur penulis mengucapkan terima kasih kepada:

1. Orang tua tercinta, papa dan mama yaitu Eldiansyah dan Noradiana yang selalu memberikan doa, dukungan dan semangat untuk penulis.
2. Atok, Memey, Dhiaz, Aunty, dan Acu yang menjadi penyemangat dan motivasi penulis dalam menyelesaikan tugas akhir.
3. Bapak Alm. Jaidan Jauhari, S.Pd., M.T. selaku Dekan Fakultas IlmuKomputer Universitas Sriwijaya.
4. Ibu Alvi Syahrini Utami, M.Kom, selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
5. Ibu Novi Yusliani, S.Kom., M.T. dan Ibu Junia Kurniati, M.Kom yang telah berkenan menjadi dosen pembimbing Tugas Akhir.
6. Bapak Dr. Abdiansah, S.Kom., M.Cs. selaku dosen pembimbing akademik atas dukungannya selama perkuliahan.
7. Mbak Wiwin Juliani selaku Admin Jurusan Teknik Informatika Bilingual dalam membantu kelancaran administrasi selama perkuliahan.

8. Hagata Aersada Ginting selaku kekasih yang memberikan inspirasi, motivasi, serta penyemangat penulis.
9. Sibul yang telah menjadi motivasi penulis dan Pawan yang telah menemani penulis selama menyelesaikan Tugas Akhir.
10. Sahabat FOVV penulis Asyraf Shafiyurrahman, Aulia Mabbruca Putri, Bintang Dwitama, Fadel Muhammad, KMH. Alviansyah, Nilam Musdalifa, Nurul Akhni, M. Raihan Almenata, Rani Sivani Yousnaidi, Reyhani Avissa, Shabrina Putri Fadhillah, dan Tarisa Rafika yang telah membantu dan menemani penulis dalam menyelesaikan Tugas Akhir.
11. Sahabat penulis Chania Ariyanti, Rifda Aliyyah, dan Mayerina Auliani Rahayu yang telah menemani serta mendukung penulis.
12. Semua pihak yang telah memberikan dukungan, semangat, dan membantu penulis dalam menyelesaikan Tugas Akhir ini.

Penulis menyadari dalam penulisan Tugas Akhir ini masih terdapat banyak kekurangan karena keterbatasan pengetahuan dan kemampuan, oleh karena itu kritik dan saran sangat diharapkan dalam membangun kemajuan penelitian selanjutnya.

Akhir kata, semoga Tugas Akhir ini dapat bermanfaat bagi kita semua.

Palembang, 22 Juli 2023

Zafira Galea

DAFTAR ISI

	Halaman
LEMBAR PENGESAHAN SKRIPSI	ii
TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI.....	iii
HALAMAN PERNYATAAN	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRACT.....	vi
ABSTRAK	vii
DAFTAR ISI.....	x
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
DAFTAR LAMPIRAN.....	xv
BAB I PENDAHULUAN	I-1
1.1 Pendahuluan	I-1
1.2 Latar Belakang	I-1
1.3 Rumusan Masalah	I-2
1.4 Tujuan Penelitian	I-3
1.5 Manfaat Penelitian	I-3
1.6 Batasan Masalah.....	I-3
1.7 Sistematika Penulisan	I-4
1.8 Kesimpulan	I-5
BAB II KAJIAN LITERATUR	II-1
2.1 Pendahuluan	II-1
2.2 Landasan Teori.....	II-1
2.2.1 Sistem Tanya Jawab	II-1
2.2.2 Pra-pengolahan	II-3
2.2.3 <i>Term Frequency - Inverse Document Frequency</i>	II-4
2.2.4 Metode Pendekatan Jarak (<i>Length Distance</i>)	II-7
2.2.5 Metode Pengujian	II-13
2.2.6 Metodologi <i>Waterfall</i> Model	II-15
2.3 Penelitian Relevan.....	II-17
2.4 Kesimpulan	II-18

BAB III METODOLOGI PENELITIAN.....	III-1
3.1 Pendahuluan	III-1
3.2 Pengumpulan Data	III-1
3.2.1 Jenis dan Sumber Data	III-1
3.2.2 Metode Pengumpulan Data	III-9
3.3 Tahap Penelitian.....	III-9
3.3.1 Pengumpulan dan Analisis Data.....	III-9
3.3.2 Membangun Sistem Tanya Jawab untuk Pertanyaan Apa Menggunakan Metode Pendekatan Jarak.....	III-10
3.3.3 Menentukan Alat Bantu Penelitian.....	III-16
3.3.4 Melakukan Pengujian	III-17
3.3.5 Menganalisis Hasil Pengujian dan Menarik Kesimpulan.....	III-18
3.3.6 Membuat Laporan Penelitian.....	III-19
3.4 Manajemen Proyek Penelitian	III-20
3.5 Kesimpulan	III-20
BAB IV PENGEMBANGAN PERANGKAT LUNAK	IV-1
4.1 Pendahuluan	IV-1
4.2 <i>Waterfall</i>	IV-1
4.2.1 <i>Requirement</i>	IV-1
4.2.2 <i>Design</i>	IV-2
4.2.3 <i>Implementation</i>	IV-9
4.2.4 <i>Verification</i>	IV-11
4.2.5 <i>Maintenance</i>	IV-11
4.3 Kesimpulan	IV-11
BAB V HASIL DAN ANALISIS PENELITIAN.....	V-1
5.1 Pendahuluan	V-1
5.2 Data Hasil Percobaan Penelitian	V-1
5.2.1 Konfigurasi Percobaan	V-1
5.2.2 Data Hasil Pengujian <i>Cosine Similarity</i>	V-4
5.2.3 Data Hasil Pengujian <i>Euclidean Distance</i>	V-8
5.2.4 Data Hasil Pengujian <i>Manhattan Distance</i>	V-12
5.3 Analisis Hasil Penelitian	V-16
5.4 Kesimpulan	V-18

BAB VI KESIMPULAN DAN SARAN	VI-1
6.1 Kesimpulan	VI-1
6.2 Saran.....	VI-2
DAFTAR PUSTAKA	xvi

DAFTAR TABEL

Tabel II-1. Klasifikasi <i>Confusion Matrix</i>	II-14
Tabel III-1. Contoh Data Pertanyaan dan Jawaban	III-2
Tabel III-2. Contoh Data Dokumen	III-3
Tabel III-3. Contoh Pertanyaan dan Jawaban	III-12
Tabel III-4. Perhitungan TF-IDF	III-12
Tabel III-5. Pengujian Pertanyaan	III-17
Tabel IV-1. Tabel Kebutuhan Fungsional	IV-2
Tabel IV-2. Tabel Kebutuhan Non-Fungsional	IV-2
Tabel IV-3. Definisi <i>Actor</i>	IV-3
Tabel IV-4. Definisi <i>Use Case</i>	IV-4
Tabel IV-5. Skenario <i>Use Case</i> Melakukan Pencarian Jawaban	IV-4
Tabel IV-6. Tabel Implementasi Objek	IV-9
Tabel IV-7. Tabel Pengujian Sistem	IV-11
Tabel V-1. Dokumen Terbaik <i>Cosine Similarity</i>	V-4
Tabel V-2. Pengujian <i>Cosine Similarity</i>	V-5
Tabel V-3. Nilai Output <i>Cosine Similarity</i>	V-5
Tabel V-4. Dokumen Terbaik <i>Euclidean Distance</i>	V-8
Tabel V-5. Pengujian <i>Euclidean Distance</i>	V-10
Tabel V-6. Nilai <i>Output Euclidean Distance</i>	V-9
Tabel V-7. Dokumen Terbaik <i>Manhattan Distance</i>	V-12
Tabel V-8. Pengujian <i>Manhattan Distance</i>	V-14
Tabel V-9. Nilai <i>Output Manhattan Distance</i>	V-13
Tabel V-10. Akurasi Algoritma dengan Skenario Terbaik	V-16

DAFTAR GAMBAR

Gambar II-1. Arsitektur Dasar Sistem Tanya Jawab (Zulqarnain et al., 2021)	II-2
Gambar II-2. <i>Measurement of text similarity</i> (Wang & Dong, 2020)	II-8
Gambar II-3. <i>Flowchart Manhattan Distance</i>	II-10
Gambar II-4. <i>Flowchart Euclidean Distance</i>	II-11
Gambar II-5. <i>Flowchart Cosine Similarity</i>	II-13
Gambar II-6. Metodologi <i>Waterfall</i> (Casteren, 2017)	II-16
Gambar III-1. Rincian Kegiatan Penelitian	III-9
Gambar III-2. Kerangka Kerja Sistem	III-10
Gambar III-3. <i>Flowchart</i> perhitungan <i>Manhattan Distance</i>	III-13
Gambar III-4. <i>Flowchart</i> perhitungan <i>Euclidean Distance</i>	III-14
Gambar III-5. <i>Flowchart</i> perhitungan <i>Cosine Similarity</i>	III-15
Gambar IV-1. Diagram <i>Use Case</i>	IV-3
Gambar IV-2. <i>Activity Diagram</i> Melakukan Pencarian Jawaban	IV-6
Gambar IV-3. <i>Sequence Diagram</i> Proses Pencarian Jawaban dengan Perhitungan Jarak Terdekat	IV-7
Gambar IV-4. <i>Class Diagram</i>	IV-8
Gambar IV-5. <i>Wireframe</i>	IV-9
Gambar IV-6. Implementasi antar muka	IV-10
Gambar V- 1. Grafik Akurasi Skenario Percobaan	V-16

DAFTAR LAMPIRAN

Lampiran 1. Kode Program	xxi
Lampiran 2. <i>User Guide</i> Program	xxii
Lampiran 3. Rencana Kegiatan Penelitian	xxiii
Lampiran 4. Sampel Data	xxv

BAB I

PENDAHULUAN

1.1 Pendahuluan

Pada bab ini akan dijelaskan tentang latar belakang penelitian Sistem Tanya Jawab menggunakan metode pendekatan jarak, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah, sistematika penulisan, serta kesimpulan penelitian yang dilakukan.

1.2 Latar Belakang

Teknologi merupakan salah satu perantara yang digunakan manusia untuk memperoleh informasi dari berbagai sumber di era digital. Sebuah informasi mengandung data yang luas serta akan terus baru dan berkembang, maka banyak sekali informasi yang dapat diperoleh. Eksplorasi data dalam jumlah besar ini membuat pencarian informasi menjadi suatu hal yang kompleks dan memakan waktu. Kesulitan ini dapat diselesaikan dengan menggunakan Sistem Tanya Jawab (Bouziane et al., 2015).

Sistem Tanya Jawab dirancang untuk membantu pengguna mencari jawaban dengan cepat dan akurat. Sistem Tanya Jawab adalah teknologi yang mengekstrak *string* dengan sistem dalam membantu menyatakan kebutuhan informasi dalam *natural language* (bahasa alami), dan secara otomatis menjawab pertanyaan dalam kutipan teks sebagai jawabannya (Hanifah & Kusumaningrum, 2021). Untuk mendapatkan pemahaman dari pertanyaan yang diajukan secara spesifik dan lebih jelas, penelitian ini akan fokus untuk pertanyaan

non-factoid “apa”. Pertanyaan “apa” mewakili pertanyaan mengenai suatu definisi, situasi, tindakan, dan alasan peristiwa.

Penelitian ini akan menemukan jawaban menggunakan metode pendekatan jarak dengan membandingkan algoritma *Euclidean Distance*, *Cosine Similarity*, dan *Manhattan Distance* dari vektor yang telah didapatkan menggunakan metode TF-IDF (*Term Frequency-Inverse Document Frequency*). Perbandingan tersebut diharapkan dapat menemukan fitur yang lebih baik dalam meningkatkan kinerja pemberian kandidat jawaban (Almiman et al., 2020). Metode pendekatan jarak memungkinkan sistem untuk mengukur akurasi dokumen yang paling relevan dengan menghitung kedekatan jarak kueri dan dokumen (Hasnat et al., 2013).

Algoritma pendekatan jarak ini dapat memberikan hasil yang akurat dalam membandingkan kesamaan antara dokumen atau kalimat. *Euclidean Distance* dan *Manhattan Distance* menghasilkan jarak yang akurat antara dua titik (Kumar Sharma & Kumar, 2016), sedangkan *Cosine Similarity* dapat mengukur kesamaan berdasarkan sudut antara dua vektor (Kitasuka et al., 2014).

1.3 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah sebagai berikut :

1. Bagaimana melakukan pencarian jawaban dari pertanyaan “apa” pada Sistem Tanya Jawab dengan mengekstraksi kata kunci menggunakan metode pendekatan jarak?
2. Bagaimana kinerja dan perbandingan algoritma dari metode pendekatan jarak *Manhattan Distance*, *Euclidean Distance*, dan

Cosine Similarity dalam pencarian jawaban pada Sistem Tanya Jawab?

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut :

1. Menghasilkan jawaban dari pertanyaan “apa” pada Sistem Tanya Jawab dengan mengekstraksi kata kunci menggunakan metode pendekatan jarak.
2. Mengetahui perbandingan algoritma menggunakan metode pendekatan jarak dalam melakukan pencarian jawaban pertanyaan “apa” pada Sistem Tanya Jawab.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut :

1. Pengguna bisa mendapatkan jawaban atas pertanyaan “apa” dengan data informasi yang berasal dari dataset.
2. Penelitian ini dapat menjadi acuan bagi penelitian berikutnya.

1.6 Batasan Masalah

Batasan masalah dari penelitian ini adalah sebagai berikut :

1. Bentuk jawaban yang diajukan dan jawaban yang diberikan hanya untuk kata “apa” tentang pertanyaan definisi dari pembelajaran biologi (*close domain*).
2. Jawaban bersumber dari dataset.
3. Data yang digunakan berbahasa Indonesia.

1.7 Sistematika Penulisan

Sistematika penulisan penelitian ini berdasarkan standar penulisan tugas akhir Fakultas Ilmu Komputer Sriwijaya, yaitu:

BAB I. PENDAHULUAN

Pada bab ini menjelaskan landasan penelitian meliputi latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, dan sistematika penulisan yang digunakan dalam menyusun laporan penelitian ini.

BAB II. KAJIAN LITERATUR

Pada bab ini memaparkan dasar-dasar teori yang digunakan dalam penelitian dan penelitian terkait yang relevan. Bab ini akan mendeskripsikan mengenai pada Sistem Tanya Jawab, metode TF-IDF sebagai representasi teks, serta metode pendekatan jarak yang dibandingkan untuk mengukur jarak teks.

BAB III. METODOLOGI PENELITIAN

Pada bab ini membahas mengenai proses tahapan dari penelitian ini yaitu, proses pengumpulan data, mendeskripsikan perencanaan penelitian dengan mengacu kepada kerangka kerja yang tersusun dan sistematis, serta perancangan manajemen proyek pada pelaksanaan penelitian.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Pada bab ini membahas mengenai pengembangan perangkat lunak berisi pembahasan mengenai perencanaan dan pembangunan perangkat lunak, mulai dari analisis kebutuhan dan desain, hingga pembuatan dan pengujian perangkat lunak.

BAB V. HASIL DAN ANALISIS PENELITIAN

Pada bab ini menguraikan hasil dan analisis penelitian yang menjadi dasar dalam penarikan kesimpulan dalam penelitian.

BAB VI. KESIMPULAN DAN SARAN

Pada bab ini akan memaparkan kesimpulan yang berisi pencapaian tujuan dari hasil penelitian dan saran yang berguna dalam penelitian selanjutnya.

1.8 Kesimpulan

Penelitian ini akan membahas mengenai pencarian jawaban untuk pertanyaan *non-factoid* dari kata “apa” pada Sistem Tanya Jawab dengan membandingkan tiga metode pendekatan jarak, yaitu *Manhattan Distance*, *Euclidean Distance*, dan *Cosine Similarity*.

BAB II

KAJIAN LITERATUR

2.1 Pendahuluan

Penelitian ini akan dijelaskan mengenai Sistem Tanya Jawab dengan membandingkan algoritma dari metode pendekatan jarak yang kuerinya telah di konversi ke dalam bentuk vektor dengan menggunakan metode TF-IDF. Pada subbab ini akan menjelaskan tentang Sistem Tanya Jawab, metode TF-IDF, serta algoritma dari metode pendekatan jarak, dengan didukung oleh penelitian yang relevan.

2.2 Landasan Teori

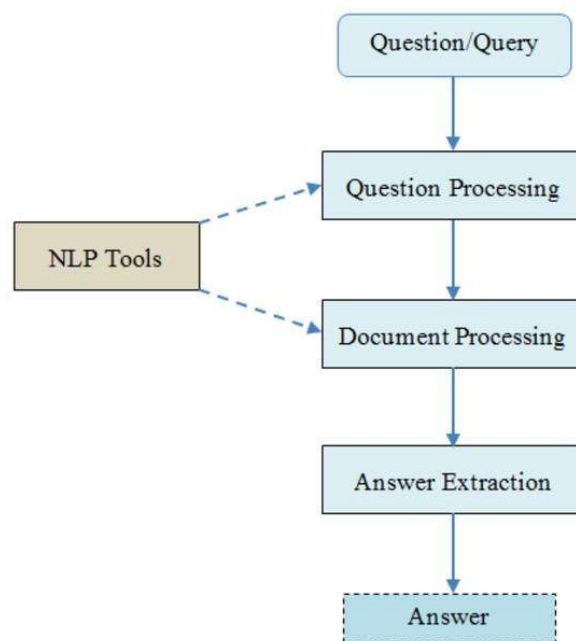
Pada subbab ini akan memuat teori-teori serta hasil penelitian yang didapatkan dari beberapa jurnal dan *paper* peneliti yang berkaitan. Tujuan ini adalah untuk mengidentifikasi hubungan dari suatu konsep serta teori dan hasil penelitian ini, akan digunakan sebagai kerangka teori dalam menyelesaikan penelitian.

2.2.1 Sistem Tanya Jawab

Sistem Tanya Jawab adalah sebuah sistem untuk menyatakan kebutuhan informasi atau pertanyaan dalam bentuk *natural language question* (pertanyaan Bahasa alami) dan lebih spesifik, serta mengembalikan kutipan teks pendek atau frasa sebagai jawabannya. Pertanyaan tersebut akan disaring untuk menentukan apakah dokumen-dokumen tersebut mengandung jawaban atas pertanyaan yang diajukan (Verberne et al., 2011).

Sistem Tanya Jawab merupakan teknologi untuk menemukan informasi dari basis teks besar menggunakan pertanyaan yang diberikan oleh pengguna (Fukumoto, 2007).

Untuk fokus pada topik pertanyaan yang dimaksud, perlu memberikan lebih banyak informasi untuk mempersempit area pencarian pertanyaan. Jika pertanyaan yang diberikan tersebut ambigu, jawabannya akan beragam sesuai pemahaman sistem dan jawaban tersebut diambil dari dokumen dengan kata-kata kueri dari kalimat tanya yang terdiri dari beberapa jenis informasi (Fukumoto et al., 2013).



Gambar II-1. Arsitektur Dasar Sistem Tanya Jawab (Zulqarnain et al., 2021)

Pada Gambar II-1, dapat dilihat Sistem Tanya Jawab dibagi atas tiga komponen yaitu *question processing*, *document processing*, dan *answer extraction*. Pada komponen *question processing* ini bertujuan untuk mendapatkan kueri dan tipe pertanyaan atau *Expected Answer Type* (EAT) terhadap pertanyaan masukan.

Kueri atau kata kunci ini digunakan dalam *document processing* dalam mencari dokumen yang memiliki kueri yang sama dengan pertanyaan masukan. Pada komponen *answer extraction*, pertanyaan dapat dibedakan sesuai jenis EAT. Proses tersebut menghasilkan *best answer* yang akan dikeluarkan oleh sistem (Yusliani & Purwarianti, 2011).

Pertama *Sistem Tanya Jawab* akan menganalisis kalimat yang diberikan untuk mendapatkan jenis pertanyaan dan menentukan kata kueri untuk pengambilan dokumen. Pertanyaan akan diekstraksi untuk mengandung kueri yang nantinya akan menjadi masukan dari *document processing* supaya dapat diklasifikasikan ke dalam kandidat kelas. Kata kueri tersebut memungkinkan ada pada dokumen yang akan diambil dengan mengenali kata kueri menggunakan modul pengenalan entitas sesuai kandidat kelasnya. Kandidat jawaban yang telah diekstraksi akan memiliki bobot dan mengurutkan kepastian jawaban yang benar. Dokumen yang terpilih adalah dokumen yang memiliki kandidat jawaban paling sering muncul dalam kata petunjuk (Fukumoto et al., 2013).

2.2.2 Pra-pengolahan

Pra-pengolahan teks adalah proses penyeleksian data agar menjadi lebih terstruktur dengan tahap *case folding*, *tokenization*, *filtering*, dan *stemming*. Tahap *case folding* yaitu mengubah semua huruf dalam dokumen menjadi huruf kecil dan menghilangkan karakter selain huruf (Yudiarta et al., 2018).

Tokenizing adalah proses penguraian dokumen atau teks menjadi unit-unit yang disebut dengan token (token ini adalah kata-kata). Lalu token yang dihasilkan

akan masuk ke proses *filtering*, di mana proses ini menghilangkan token berdasarkan *stopword* dan membuang karakter-karakter yang tidak digunakan seperti tanda baca dan *white space* agar tidak mengganggu proses pencarian data (Ryansyah & Andayani, 2017). Selanjutnya melalui tahap *stemming* yaitu pencarian kata dasar dari setiap kata dengan menghilangkan imbuhan (Yudiarta et al., 2018). Setelah kedua proses tersebut token akan dibobot agar dapat menghitung kesamaan dokumen atau teks (Ryansyah & Andayani, 2017) dari kandidat jawaban.

2.2.3 Term Frequency - Inverse Document Frequency

Term Frequency - Inverse Document (TF-IDF) adalah penilaian komprehensif tentang pentingnya sebuah kata untuk sebuah teks atau dokumen dan juga sebagai representasi sebuah teks. Nilai *Term Frequency* yang relatif besar akan memengaruhi keakuratan nilai TF-IDF dalam menilai kontribusi kata dalam klasifikasi, serta meningkatkan nilai komputasi. Nilai TF-IDF. Nilai TF-IDF yang dihitung merupakan kata yang muncul di paragraf pertama serta kata pada teks dokumen dan kata benda yang masing-masing telah dikalikan dan menghasilkan bobot (Jones & Sparck, 1972).

$$tf - idf_{(t,d)} = tf_{(t,d)} \times idf_{(t)} \quad (\text{II} - 1)$$

Keterangan:

$tf - idf_{(t,d)}$ = bobot *term*

$tf_{(t,d)}$ = *term frequency* kata (t) pada dokumen (d)

$idf_{(t)}$ = *invers document frequency* kata (t)

Pengambilan kueri yang diberikan pengguna akan diubah menjadi bentuk vektor untuk memisahkan kesamaan semantik secara efektif. Hal tersebut dapat dilakukan dengan menggunakan metode TF-IDF yang berperan dalam mengonversi kueri ke bentuk vektornya. Kueri vektor yang dihasilkan oleh TF-IDF ini sangat penting untuk menunjukkan kesamaan tekstual tertentu dalam memisahkan informasi kontekstual yang berbeda dari pertanyaan yang sama (Patil & Tiwari, 2020).

2.2.3.1 *Term Frequency (TF)*

Term Frequency (TF) adalah frekuensi kemunculan kata di dalam sebuah dokumen dan pertanyaan, yang secara intuitif menunjukkan pentingnya kata dalam dokumen dan pertanyaan tersebut (Liang & Niu, 2022).

Term Frequency (TF) mengandung sebuah bobot dari suatu kata (t) yang dihitung dari sebuah dokumen (d) dan dilambangkan dengan $tf_{t,d}$. Pendekatan konsep *term frequency* ini yaitu dengan menyatakan bobot suatu kata (t) sebagai jumlah kemunculan pada dokumen (d) tersebut (Ryansyah & Andayani, 2017).

$$tf_{(t,d)} = \frac{\text{number of times } t \text{ appears in } d}{\text{total number of terms in } d} \quad (\text{II} - 2)$$

Keterangan:

$tf_{(t,d)}$ = *term frequency* kata (t) pada dokumen (d)

Kelemahan *term frequency* adalah menganggap semua kata setara, yang mengakibatkan relevansi suatu kata tinggi jika kata tersebut sering muncul dalam dokumen-dokumen tersebut. Padahal tingginya frekuensi kemunculan suatu kata

belum menentukan kata tersebut merupakan kata yang penting (Ryansyah & Andayani, 2017).

2.2.3.2 *Inverse Document Frequency (IDF)*

Invers Document Frequency (IDF) berguna dalam mengukur pentingnya suatu istilah. *Invers Document Frequency* dibutuhkan untuk mengurangi istilah yang sering muncul sambil meningkatkan istilah yang jarang. Hal ini dilakukan karena saat menghitung *term frequency* (TF), semua istilah dianggap sama pentingnya. Namun beberapa istilah yang muncul berkali-kali belum menentukan pentingnya suatu istilah tersebut, seperti “adalah”, “dari”, dan “itu” (Patil & Tiwari, 2020).

$$idf_{(t)} = \log \frac{N}{N_{(t)}} \quad (\text{II} - 3)$$

Keterangan:

$idf_{(t,d)}$ = bobot frekuensi kolektif

N = jumlah keseluruhan dokumen

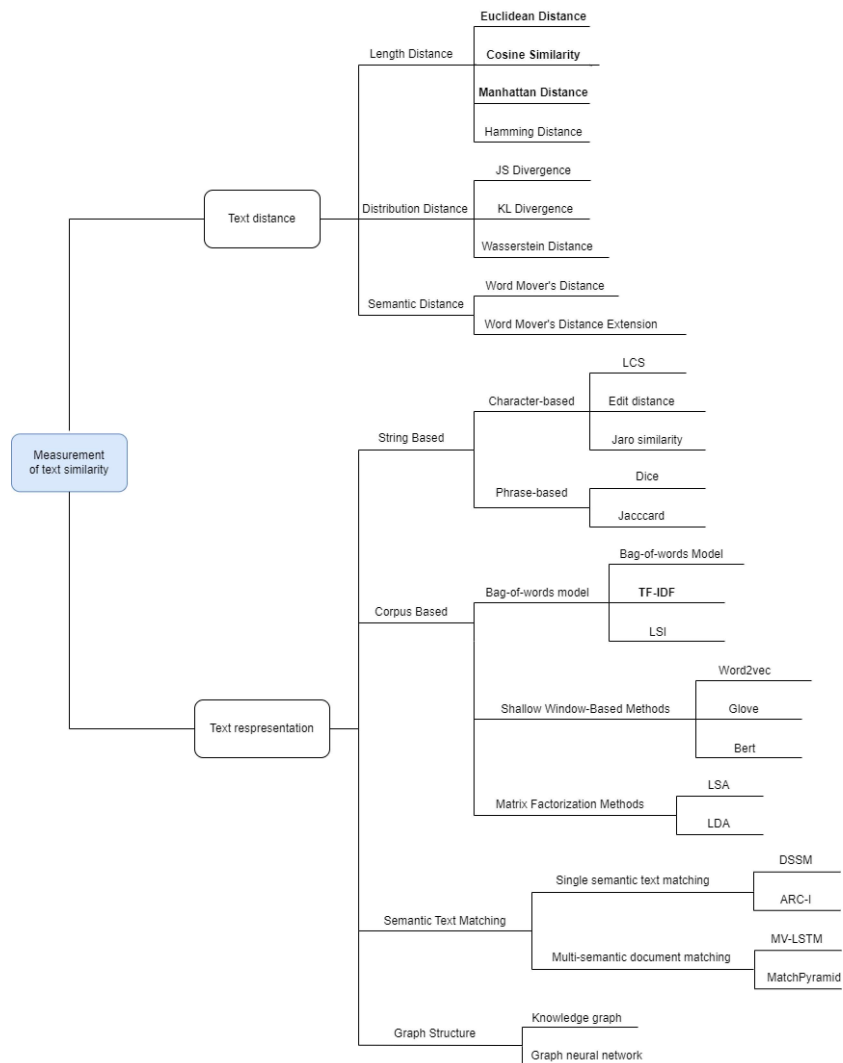
$N_{(t)}$ = jumlah dokumen yang memuat *term* (t)

Tujuan *Invers Document Frequency* (IDF) yaitu menurunkan bobot dari kata dengan frekuensi kolektif (frekuensi total kemunculan kata di semua dokumen) yang bernilai tinggi. Dapat disebut juga, semakin banyak dokumen kata tersebut

pada suatu kumpulan dokumen, maka semakin rendah bobotnya (Ryansyah & Andayani, 2017).

2.2.4 Metode Pendekatan Jarak (*Length Distance*)

Dalam pengukuran kemiripan teks dijelaskan dengan dua aspek yaitu representasi teks dan jarak teks. Jarak teks sendiri dibagi menjadi tiga, mengukur pendekatan jarak, jarak distribusi, dan jarak simantik. Pendekatan jarak sendiri dapat dihitung dengan mengukur *Euclidean Distance*, *Cosine Similarity*, *Manhattan Distance*, atau *Hamming Distance* (Wang & Dong, 2020). Pembagian aspek serta metode dapat dilihat pada Gambar II-2.



Gambar II-2. *Measurement of text similarity* (Wang & Dong, 2020)

Salah satu cara untuk menghitung bobot dalam pengurutan kepastian jawaban yang benar dalam Sistem Tanya Jawab adalah dengan cara menghitung bobot jarak kata menggunakan kata-kata kueri (Fukumoto et al., 2013).

2.2.4.1 *Manhattan Distance*

Manhattan Distance atau juga disebut dengan jarak blok, jarak nilai absolut, maupun jarak gerbong ini adalah jarak yang dihitung sesuai dengan jarak yang akan

ditempuh untuk mendapatkan dari satu titik data ke yang lain jika mengikuti jalur *grid* (Gomaa & Fahmy, 2013), tapi bukan jarak geometris. Perhitungan ini umumnya hanya berfungsi jika titik-titik tersebut disusun dalam bentuk kisi-kisi (Wang & Dong, 2020).

$$d(S_a, S_b) = \sum_{i=1}^n |S_a^{(i)} - S_b^{(i)}| \quad (\text{II} - 4)$$

Keterangan :

d = *distance*

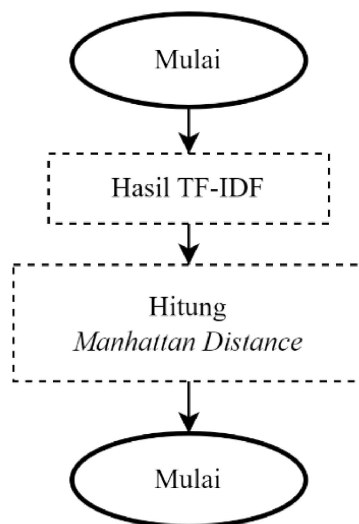
n = jumlah observasi

S_a = vektor A yang dibentuk dari dokumen A

S_b = vektor B yang dibentuk dari dokumen B

$S_a^{(i)}$ = nilai sumbu horizontal pada bidang koordinat

$S_b^{(i)}$ = nilai sumbu vertikal pada bidang koordinat



Gambar II-3. *Flowchart Manhattan Distance*

Pada Gambar II-3, dapat dilihat *flowchart* dari algoritma *Manhattan Distance*. Setelah mendapatkan hasil perhitungan TF-IDF, pertanyaan dan kandidat jawaban akan dihitung serta dibandingkan hasil kemiripannya. Algoritma *Manhattan Distance* ini cenderung lebih cepat dibandingkan *Euclidian Distance* pada perhitungan manual karena jumlah iterasi pada algoritma ini cenderung lebih sedikit. *Manhattan Distance* memiliki tingkat akurasi yang lebih lemah (Pribadi et al., 2022).

2.2.4.2 Euclidean Distance

Euclidean Distance merupakan akar kuadrat dari jumlah perbedaan kuadrat antara elemen yang sesuai dari dua vektor (Gomaa & Fahmy, 2013). Jarak perhitungan ini mengambil jarak garis lurus antara dua titik (Wang & Dong, 2020).

$$d(S_a, S_b) = \sqrt{\sum_{i=1}^n (S_a^{(i)} - S_b^{(i)})^2} \quad (\text{II} - 5)$$

Keterangan:

d = distance

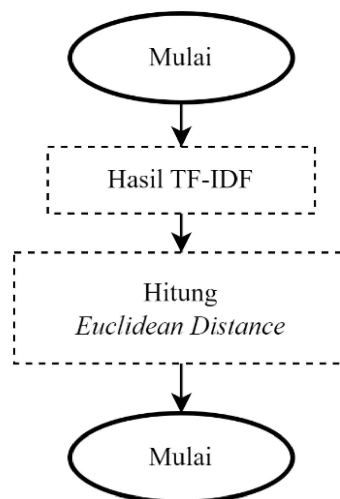
n = jumlah observasi

S_a = vektor A yang dibentuk dari dokumen A

S_b = vektor B yang dibentuk dari dokumen B

$S_a^{(i)}$ = nilai sumbu horizontal pada bidang koordinat

$S_b^{(i)}$ = nilai sumbu vertikal pada bidang koordinat



Gambar II-4. *Flowchart Euclidean Distance*

Pada Gambar II-4, dapat dilihat *flowchart* dari algoritma *Euclidean Distance*. Setelah mendapatkan hasil perhitungan TF-IDF, pertanyaan dan kandidat

jawaban akan dihitung serta dibandingkan hasil kemiripannya. Algoritma *Euclidean Distance* memiliki kelebihan dan kekurangan menurut (Pribadi et al., 2022), kelebihannya adalah dapat memiliki hasil yang lebih optimal dibandingkan dengan perhitungan yang lainnya. Kekurangannya adalah perhitungan manual dengan algoritma ini cenderung lebih lama karena jumlah iterasi *Euclidean Distance* lebih banyak dibandingkan dengan *Manhattan Distance*.

2.2.4.3 Cosine Similarity

Cosine Similarity adalah ukuran kesamaan antara dua vektor ruang hasil kali dalam yang mengukur sudut *cosinus* di antara vektor tersebut (Gomaa & Fahmy, 2013). Nilai *cosine* memiliki rentang dari nol hingga satu dengan hasil nilai yang semakin besar dapat diartikan tingkat kesamaan yang semakin tinggi juga (Ryansyah & Andayani, 2017).

$$sim = \frac{\vec{A} \cdot \vec{B}}{\|A\| \cdot \|B\|} \quad (\text{II} - 6)$$

Keterangan:

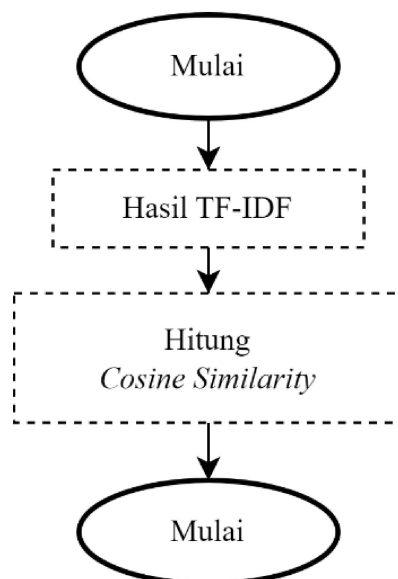
sim = tingkat kesamaan

\vec{A} = vektor A yang dibentuk dari dokumen A

\vec{B} = vektor B yang dibentuk dari dokumen B

$\|A\|$ = Panjang vektor A

$\|B\|$ = Panjang vektor



Gambar II-5. *Flowchart Cosine Similarity*

Pada Gambar II-5, dapat dilihat *flowchart* dari algoritma *Cosine Similarity*. Setelah mendapatkan hasil perhitungan TF-IDF, pertanyaan dan kandidat jawaban akan dihitung serta dibandingkan hasil kemiripannya. Kelebihan dari algoritma *Cosine Similarity* adalah tidak terpengaruh pada panjang pendeknya suatu dokumen dan memiliki tingkat akurasi yang tinggi. Sedangkan pada penelitian (Firdaus, et al., 2014), mendapatkan kelemahan *Cosine Similarity* yaitu tidak dapat menentukan sinonim dari kata yang dibandingkan.

2.2.5 Metode Pengujian

Confusion matrix merupakan ukuran penting untuk mengevaluasi akurasi model klasifikasi (Zeng, 2020). *Confusion matrix* berisi mengenai informasi tentang klasifikasi aktual dan prediksi yang dilakukan oleh sistem klasifikasi dengan menggunakan data matriks (Santra & Christy, 2012). Tabel II-1, menunjukkan pengklasifikasian dari *confusion matrix*.

Tabel II-1. Klasifikasi *Confusion Matrix*

	Predicted Values		
Actual Size		Positive	Negative
	Positive	TP	FN
	Negative	FP	TN

Terdapat 4 istilah yang digunakan dalam proses klasifikasi pada *confusion matrix*.

1. TP (*True Positive*) merupakan data jawaban benar keluaran sistem yang sesuai dengan fakta diklasifikasikan sebagai *positive*.
2. FN (*False Negative*) merupakan data salah keluaran sistem yang tidak sesuai dengan fakta diklasifikasikan sebagai *negative*.
3. FP (*False Positive*) merupakan data salah keluaran sistem yang tidak sesuai dengan fakta diklasifikasikan sebagai *positive*.
4. TN (*True Negative*) merupakan data benar keluaran sistem yang sesuai dengan fakta diklasifikasikan dengan *negative*.

Confusion matrix dapat mengukur nilai *accuracy*, *precision*, dan *recall*. *Accuracy* adalah proporsi jumlah total prediksi yang benar untuk mempresentasikan seberapa tepat jawaban yang diberikan oleh sistem berdasarkan pertanyaan yang diajukan. *Precision* adalah proporsi prediksi benar *positive* yang dibandingkan dengan jumlah klasifikasi nilai *positive*. *Recall* adalah proporsi nilai *positive* yang terprediksi benar.

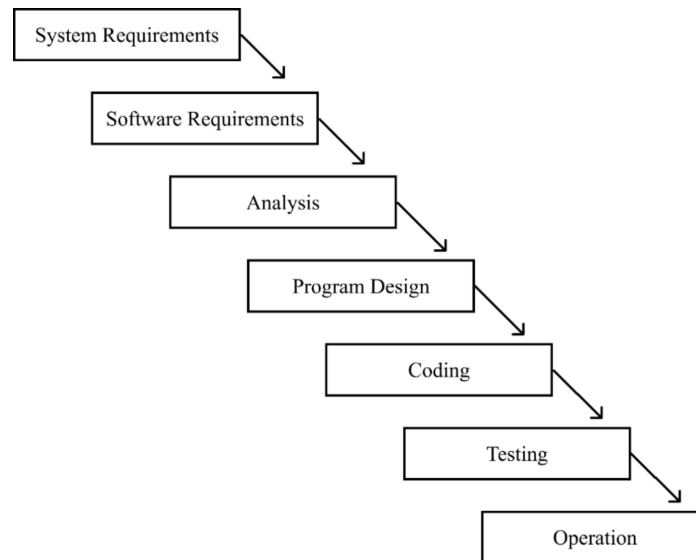
$$Accuracy = \frac{TP+TN}{TP+FP+FN+FN} \times 100\% \quad (II-7)$$

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (II-8)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (II-9)$$

2.2.6 Metodologi *Waterfall Model*

Model *waterfall* merupakan salah satu metodologi SDLC (*System Development Life Cycle*) dalam mengembangkan sistem informasi atau perangkat lunak. Model ini menggunakan pendekatan sistematis dan berurutan dari tahap perencanaan hingga tahap pengelolaan (*maintenance*) (Wahid, 2020). Model ini menekankan ketergantungan pada tahap sebelumnya yang harus dilakukan terlebih dahulu untuk menangani tahap selanjutnya (Casteren, 2017). Metodologi ini memiliki kelebihan menghasilkan kualitas sistem yang baik karena pelaksanaan pengembangan sistem informasi ini dilakukan secara bertahap. *Waterfall* cocok digunakan untuk proyek pembuatan sistem baru dan pengembangan sistem atau perangkat lunak yang berskala besar (Wahid, 2020).



Gambar II-6. Metodologi *Waterfall* (Casteren, 2017)

Pada Gambar II-6, dapat dilihat tahapan metodologi perangkat lunak.

Tahapan metodologi ini dijelaskan sebagai berikut:

1. *Requirement* dan *Analysis* merupakan tahap dalam memahami perangkat lunak yang diharapkan oleh pengguna serta batasan perangkat lunak tersebut.
2. *Program Design* merupakan tahap merancang arsitektur sistem secara keseluruhan sebagai penentu perangkat keras (*hardware*).
3. *Coding* atau *Implementation* adalah tahap pertama kali dikembangkan di program kecil (*unit*).
4. *Testing* atau *Verification*, tahap ini menguji sistem apakah sudah sesuai dengan persyaratan sistem.
5. *Operation* atau *Maintenance* merupakan tahap akhir dari pengembangan perangkat lunak, yaitu dengan melakukan pemeliharaan sebuah perangkat

lunak tersebut dengan memperbaiki kesalahan yang tidak ditemukan sebelumnya.

2.3 Penelitian Relevan

Pada subbab ini memaparkan beberapa penelitian terkait mengenai Sistem Tanya Jawab.

Pada penelitian terkait “A Stacked BiLSTM Neural Network Based on Coattention Mechanism for Question Answering” yang diteliti oleh (Cai et al., 2019) membahas mengenai BiLSTM deep learnig model yang menggabungkan mekanisme atensi secara simantik dalam memahami dan memodelkan question answering, dengan memetakan pertanyaan dan jawaban ke vektor distribusi yang sesuai. Lalu menghitung jarak serta kesamaan simantik dengan Cosinus Similarity dan Euclidean Distance. Hal ini memungkinkan model untuk menangkap fitur sentence-level ketergantungan yang panjang dan menghasilkan representasi codependent yang lebih dalam untuk pasangan question answering. Hasil MAP dan MRR terbaik yang dicapai dengan metode ini yaitu, 0,7613 dan 0,8401.

Penelitian lain yang dilakukan oleh (Ryansyah & Andayani, 2017) dengan judul “Implementasi Algoritma TF-IDF Pada Pengukuran Kesamaan Dokumen” membahas mengenai pengukuran antara dokumen dan mengukur tingkat kesamaan dokumen. Kesamaan dokumen ini diukur dengan menggunakan metode TF-IDF untuk menentukan bobot dan menghasilkan tingkat kemiripan antara dokumennya dengan algoritma *Cosine Similarity*. Hasil dari pengujian algoritma ini membutuhkan minimal tiga dokumen serta hasil uji korelasi format dokumen ini mempengaruhi jumlah karakter dan waktu pemrosesan dokumen.

Penelitian terkait lainnya telah dilakukan oleh (Hanifah & Kusumaningrum, 2021) yang berjudul “Non-Factoid Answer Selection in Indonesian Science Question Answering System using Long Short-Term Memory (LSTM)” dengan memilih jawaban *non-factoid* untuk Sistem Tanya Jawab yang dilakukannya. Penelitian ini menggabungkan model Word2Vec dan model LSTM, serta menggunakan algoritma *cosinus similarity* untuk menentukan relevansi pertanyaan dan jawaban. Sebanyak 400 pasang data tentang soal dan jawaban, disimpulkan model LSTM berhasil dalam pemilihan jawaban pertanyaan *non-factoid* dengan nilai MRR yang dicapai sebesar 90.06%, dan nilai MAP sebesar 78,69%.

Penelitian yang telah dilakukan oleh (Li et al., 2022) dengan judul “Towards Knowledge-Based Tourism Chinese Question Answering System” membahas Sistem Tanya Jawab untuk membangun grafik pengetahuan yang lebih komprehensif dengan hasil memasukan *embedding* model BERT ke dalam pengklasifikasi. Lalu dengan algoritma *Manhattan Distance* dalam mengklasifikasikan label kelas (mencari kata kunci) pertanyaan. Jawaban akhir didapatkan setelah mencari grafik pengetahuan dengan menggabungkannya. Data yang digunakan diperoleh dari situs web pariwisata dan data kata kunci Baidu di beberapa tempat. Dengan memperoleh hasil akurasi yang tinggi dalam menjawab pertanyaan sederhana dan akurasinya ditingkatkan dalam menjawab pertanyaan kompleks.

2.4 Kesimpulan

Pada bab ini telah dipaparkan landasan teori terkait penelitian ini, yaitu mengenai Sistem Tanya Jawab, TF-IDF, metode pendekatan jarak, serta penelitian

lain yang relevan dengan penelitian ini. Dalam penelitian ini membandingkan tiga metode pendekatan jarak dengan cara perhitungan yang berbeda seperti yang telah dideskripsikan pada bab ini.

BAB III

METODOLOGI PENELITIAN

3.1 Pendahuluan

Pada bab ini akan membahas mengenai proses serta tahapan dari pengumpulan data yang digunakan dan akan dibahas secara lengkap dan sistematis. Adapun tujuan dari bab ini adalah sebagai kumpulan data dalam mengembangkan penelitian dan agar peneliti memahami alur dan langkah-langkah dalam melakukan penelitian.

3.2 Pengumpulan Data

Pada subbab ini akan dipaparkan mengenai jenis dan sumber data serta metode pengumpulan data yang dilakukan pada penelitian ini.

3.2.1 Jenis dan Sumber Data

Jenis data yang digunakan dalam penelitian ini adalah data sekunder. Penelitian ini menggunakan dataset sebanyak 110 pasang data pertanyaan beserta jawabannya dan dokumen sebanyak 61 dokumen yang berbentuk .txt. Dataset ini bersumber dari jurnal ‘Sistem *Question Answering* Bahasa Indonesia untuk Pertanyaan *Non-Factoid* oleh (Yusliani & Purwarianti, 2011) tentang pertanyaan definisi dari pembelajaran biologi. Adapun beberapa contoh data pertanyaan dan jawaban yang dapat dilihat pada Tabel III-1. Pada Tabel III-2 merupakan tabel mengenai contoh salah satu data dokumen.

Tabel III-1. Contoh Data Pertanyaan dan Jawaban

Pertanyaan	Jawaban
Apakah yang dimaksud dengan SEL ?	[sel dapat diartikan sebagai unit terkecil penyusun makhluk hidup dan sebagai tempat berlangsungnya aktivitas kehidupan.]
Apakah yang dimaksud dengan Protoplasma ?	[Protoplasma adalah sejenis substansi kompleks seperti agar-agar yang tidak habis digunakan saat aktivitas kimiawi dalam menjaga kelangsungan hidup sel.]
Apakah yang dimaksud dengan sitoplasma ?	[Sitoplasma merupakan cairan yang mengelilingi inti sel dengan membran sel sebagai batas luarnya. Dasar penyusunnya ialah sitosol yang bersifat koloid.]
Apakah yang dimaksud dengan Prokariotik ?	[Prokariotik (prokaryote) berasal dari bahasa Yunani, yakni pro artinya 'sebelum' dan karyon artinya 'kernel' atau 'nukleus'. Berdasarkan asal kata tersebut, sel prokariotik diartikan sebagai sel makhluk hidup yang tidak bernukleus.]
Apakah yang dimaksud dengan Sel Eukariotik ?	[Sel eukariotik (Yunani: eu, berarti sebenarnya) merupakan sel makhluk hidup bernukleus yang diselaputi membran. Di dalam membran ini terdapat cairan yang disebut sitoplasma.]

Tabel III-2. Contoh Data Dokumen

Nama Dokumen	Isi Dokumen
1.txt	<p><j>Alat Pernapasan. <p>Bernapas adalah kegiatan menghirup udara dan mengeluarkan udara. Udara mengandung berbagai komponen gas, salah satunya adalah oksigen (O₂). Oksigen inilah yang diperlukan oleh tubuh. Oksigen masuk ke dalam tubuh melalui pernapasan. Selanjutnya, pernapasan menghasilkan karbon dioksida (CO₂) yang dikeluarkan dari dalam tubuh. Bernapas menggunakan alat-alat pernapasan. Alat pernapasan manusia terdiri atas hidung, tenggorokan, dan paru-paru.</p></p> <p><p>Hidung merupakan tempat keluar masuknya udara pernapasan. Udara masuk melalui lubang hidung menuju rongga hidung. Di dalam rongga hidung terdapat rambut hidung dan selaput lendir. Rambut hidung dan selaput lendir berfungsi menyaring udara yang masuk agar bebas dari debu dan kuman. Dengan demikian, udara yang kita hirup bersih dari kotoran, debu, maupun kuman penyakit. Di dalam hidung udara juga mengalami penyesuaian suhu dan kelembapan.</p></p> <p><p>Udara pernapasan dari hidung turun ke tenggorokan (trakea). Tenggorokan merupakan sebuah saluran yang panjangnya kira-kira 9 cm. Pada tenggorokan terdapat bulu-bulu halus. Bulu-bulu halus berfungsi menyaring udara dari kotoran yang masih dapat lolos ke tenggorokan. Ujung trakea bercabang menjadi dua bagian. Cabang-cabang ini disebut Bronkus kanan menuju paru-paru kanan. Bronkus kiri menuju paru-paru kiri.</p></p>

	<p> <p>Paru-paru terdapat di dalam rongga dada di atas diafragma. Diafragma adalah sekat antara rongga dada dan rongga perut. Paru-paru ada dua buah yaitu paru-paru kiri dan paru-paru kanan. Paru-paru kiri terdiri atas dua gelambir. Paru-paru kanan terdiri atas tiga gelambir. Paru-paru dibungkus oleh selaput paru-paru yang disebut pleura. Di dalam paru-paru terdapat cabang-cabang bronkus yang disebut bronkiolus. Bronkiolus juga memiliki percabangan yang jumlahnya sangat banyak. Cabang-cabang tersebut sangat halus dan tipis. Tiap-tiap ujung cabang membentuk kantung berdinding tipis yang disebut alveolus. Alveolus merupakan gelembung yang sangat tipis. Gelembung tersebut diselimuti pembuluh kapiler darah. Pada alveolus terjadi pertukaran gas O₂ dan CO₂. Pada saat udara yang kita hirup sampai di alveolus, oksigen melewati dinding kapiler darah. Oksigen diikat oleh hemoglobin (Hb) darah. Setelah itu, darah akan mengedarkan oksigen ke seluruh tubuh. Dalam tubuh, oksigen digunakan untuk proses pembentukan energi. Pada proses tersebut dihasilkan energi dan gas karbon dioksida (CO₂). CO₂ tersebut diikat kembali oleh hemoglobin darah. Setelah itu, darah akan membawa CO₂ ke paru-paru. CO₂ dari paru-paru menuju tenggorokan, kemudian ke lubang hidung untuk dikeluarkan dari dalam tubuh.</p> </p> <p> <p>Masuknya O₂ dan keluarnya CO₂ pada saluran pernapasan terjadi pada saat berlangsungnya proses pernapasan. Proses-proses ini diatur oleh otot diafragma dan otot di antara tulang rusuk. Pada saat menarik napas otot diafragma mengerut. </p> </p>
--	---

<p>Akibatnya, diafragma mendatar, rongga dada membesar, dan udara masuk paru-paru. Selain itu, paru-paru dapat pula terisi udara dengan mengerutnya otot antar tulang rusuk. Otot antar tulang rusuk yang mengerut menyebabkan rongga dada membesar dan udara masuk ke dalam paru-paru. Proses masuknya udara pernapasan ke dalam paru-paru disebut inspirasi.</p> <p>Pada saat mengembuskan napas, otot diafragma dan otot antar tulang rusuk mengendur. Akibatnya, rongga dada mengecil dan paru-paru mengempis sehingga CO₂ dalam paru-paru terdorong keluar. Proses tersebut merupakan proses ekspirasi.</p> <p><p>Proses pernapasan dapat terganggu jika ada salah satu alat pernapasan mengalami gangguan. Gangguan tersebut dapat disebabkan oleh kuman maupun polusi udara. Beberapa gangguan maupun penyakit pada alat pernapasan sebagai berikut.</p> <p>a) Influenza (flu) merupakan penyakit yang disebabkan oleh virus. Orang yang terserang flu akan mengalami demam, menggigil, batuk, sakit kepala, bersin-bersin, serta nyeri punggung. Lendir yang keluar dari hidung menutup lubang hidung sehingga udara terhalang masuk dan mengganggu pernapasan.</p> <p>b) Sesak napas merupakan gangguan pernapasan karena udara yang tercemar oleh asap. Asap dapat berasal dari pembakaran sampah, kendaraan bermotor, dan rokok. Selain asap, debu juga dapat mengakibatkan sesak napas.</p> <p>c) Asma yaitu gangguan pernapasan karena penyempitan saluran pernapasan. Menyempitnya saluran pernapasan dapat terjadi karena beberapa hal berikut.</p> <p>1) Udara yang tercemar oleh asap dan debu.</p>	
---	--

	<p>2) Udara yang terlalu dingin.</p> <p>3) Keadaan jiwa penderita, misalnya stres dan tekanan emosi.</p> <p>d) Radang paru-paru karena bakteri Tuberkulosis. Radang yang disebabkan oleh bakteri ini biasa disebut TBC paru-paru.</p> <p>e) Bronkitis yaitu adanya peradangan pada batang tenggorok (bronkus).</p> <p>f) Polip merupakan penyempitan saluran pernapasan akibat terjadinya pembengkakan kelenjar limfe.</p> <p><p>Gangguan pada alat-alat pernapasan dapat mengganggu aktivitas sehari-hari. Oleh karena itu, jagalah kesehatan alat pernapasanmu dengan membiasakan diri berpola hidup sehat!. Pola hidup sehat tersebut di antaranya sebagai berikut.</p> <ol style="list-style-type: none"> Berolahraga secara teratur. Menjaga sirkulasi udara di rumah. Mengonsumsi makanan sehat dan bergizi seimbang. Istirahat teratur. Mengenakan masker saat berkendara. Tidak merokok. <p></p></p> <p><p>Hewan juga bernapas seperti halnya manusia. Hewan tertentu memiliki alat pernapasan khusus sesuai tempat hidupnya. Hewan-hewan yang memiliki alat pernapasan khusus yaitu ikan dan cacing tanah.</p></p> <p><p>Ikan memiliki alat pernapasan berupa insang. Insang terletak di sebelah kanan dan kiri kepala. Insang ini berjumlah empat pasang. Bagian-bagian insang berbentuk lembaran yang disebut lembaran insang. Pada lembaran insang terjadi</p>
--	--

<p>pertukaran udara. Ikan juga mempunyai gelembung renang untuk menyimpan oksigen dan mengatur gerak. Ikan memperoleh oksigen dari dalam air. Mekanisme pernapasan ikan melalui beberapa tahap. Mula-mula ikan membuka mulutnya untuk mengambil air. Air kemudian mengalir masuk ke rongga mulut menuju lembaran insang. Setelah itu, air keluar melewati tutup insang. Ketika air melewati lembaran insang, oksigen diikat oleh Hb (hemoglobin) darah. Pada saat yang sama, Hb juga melepaskan karbon dioksida ke air. Apakah Labirin Itu?. Labirin berupa lipatan-lipatan tidak teratur yang merupakan perluasan rongga insang. Lipatan-lipatan itu terletak di atas rongga insang. Labirin berguna untuk menyimpan udara. Bagi ikan-ikan yang hidup di air keruh atau di rawa-rawa, labirin sangat membantu untuk bernapas. Di tempat-tempat tersebut kandungan oksigennya kurang. Oleh karena itu, ikan sering menuju permukaan air untuk mengambil oksigen dari udara. Oksigen tersebut disimpan dalam labirin. Ikan yang memiliki labirin, misalnya ikan gabus dan lele.</p><p>Cacing tanah bernapas melalui permukaan tubuhnya. Cacing tanah memiliki kulit yang tipis. Pada permukaan kulit cacing tanah terdapat banyak pembuluh darah. Kulit cacing tanah juga menghasilkan lendir. Oleh karena itu, kulit terlihat basah dan lembap. Kondisi ini menyebabkan cacing dapat menyerap oksigen dan mengeluarkan karbon dioksida secara langsung melalui permukaan kulit. Itulah sebabnya cacing membutuhkan tempat lembap atau basah. Tempat lembap membantu proses pernapasan agar dapat berlangsung dengan baik.</p></p>	
---	--

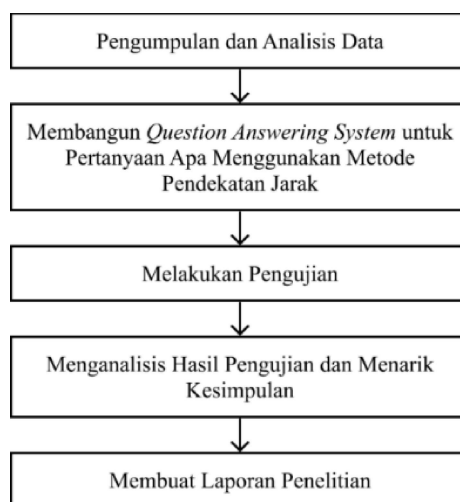
	<p><p>Burung bernapas dengan paru-paru. Selain paru-paru, pernapasan pada burung juga dibantu oleh pundi-pundi (kantong) udara. Pundi-pundi udara ini merupakan alat bantu pernapasan, terutama pada saat terbang. Pada saat terbang, burung menyimpan udara di dalam pundi-pundi tersebut. Pada saat burung tidak terbang, pernapasannya dilakukan dengan cara menghirup udara melalui hidung, tenggorokan, paru-paru, dan pundi-pundi udara. Pada paru-paru inilah terjadi pengikatan oksigen dan pelepasan karbon dioksida serta uap air. Pada saat terbang, burung bernapas dengan cara mengalirkan udara yang ada di dalam pundi-pundi udara melalui gerakan sayapnya. Gerakan kedua sayapnya inilah yang menyebabkan pundi-pundi udara mengembang dan mengempis sehingga udara dapat masuk ke dalam paru-paru.</p></p> <p><p>Untuk melakukan proses pernapasan, serangga menggunakan trakea sebagai alat pernapasannya. Trakea merupakan pembuluh-pembuluh halus yang bercabang-cabang dan tersebar ke seluruh tubuh. Pada ujung pembuluh ini terdapat lubang-lubang pernapasan yang disebut stigma. Stigma terletak di sepanjang kedua sisi tubuh serangga dan berfungsi sebagai jalan keluar dan masuknya udara. Jadi, pada serangga pernapasan dimulai dengan masuknya udara melalui stigma, kemudian udara tersebut dialirkan ke seluruh tubuh oleh trakea.</p></p>
--	---

3.2.2 Metode Pengumpulan Data

Metode pengumpulan data, dilakukan dengan mengumpulkan data sekunder yang diperoleh dari jurnal ‘Sistem *Question Answering* Bahasa Indonesia untuk Pertanyaan *Non-Factoid*’ oleh (Yusliani & Purwarianti, 2011) tentang pertanyaan definisi dari pembelajaran biologi berbentuk .txt.

3.3 Tahap Penelitian

Tahapan penelitian merupakan langkah-langkah aktivitas penelitian secara terstruktur dan sistematis dalam kegiatan penelitian yang dilakukan. Adapun rincian kegiatan dapat dilihat pada Gambar III-1.



Gambar III-1. Rincian Kegiatan Penelitian

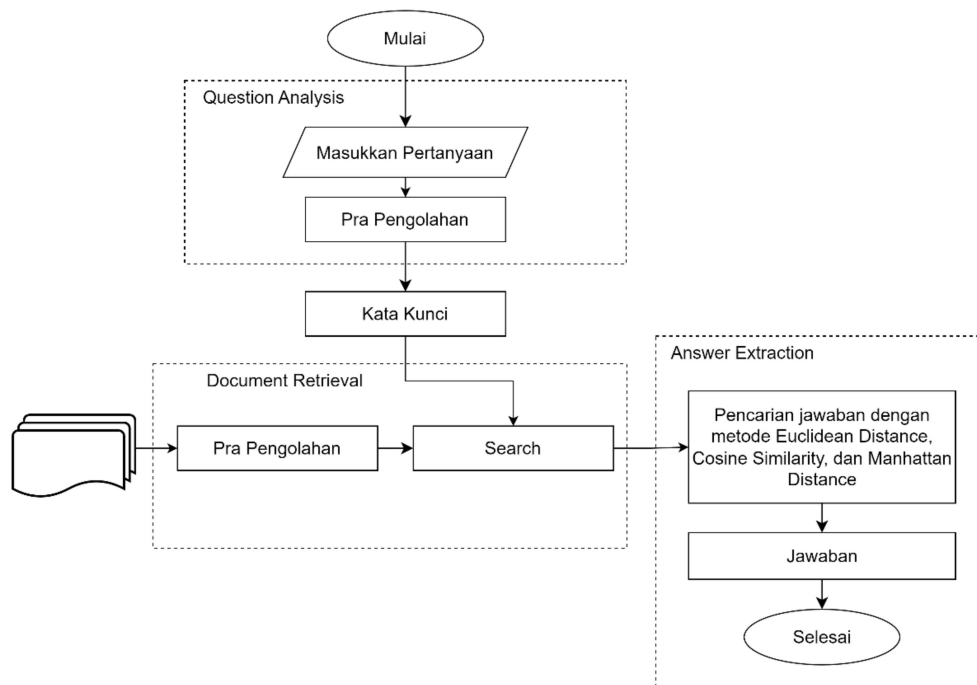
3.3.1 Pengumpulan dan Analisis Data

Pada tahapan mengumpulkan data, akan dilakukan proses pengumpulan data sekunder dengan 110 pasang data pertanyaan dan jawaban serta 61 dokumen yang mengandung jawaban. Data yang diambil berdasarkan informasi yang mengandung pertanyaan definisi “apa” dan akan dilakukan pra-pengolahan data.

3.3.2 Membangun Sistem Tanya Jawab untuk Pertanyaan Apa

Menggunakan Metode Pendekatan Jarak

Penelitian ini membangun Sistem Tanya Jawab untuk pertanyaan definisi ‘apa’ menggunakan metode pendekatan jarak dengan membandingkan algoritma *Manhattan Distance*, *Euclidean Distance*, dan *Cosine Similarity*. Algoritma yang dibandingkan dalam metode ini, untuk menghasilkan penggunaan metode yang lebih efisien dan baik. Penelitian ini akan dibangun dengan kerangka kerja sistem pada Gambar III-2.



Gambar III-2. Kerangka Kerja Sistem

Berdasarkan Gambar III-2, proses sistem yang dilakukan pada penelitian ini melalui tahap:

1. Memasukkan pertanyaan apa *non-factoid* pada *website*.

2. Pra-pengolahan, pada tahap *question analysis* ini data pertanyaan akan melewati proses *case folding* dengan mengubah karakter menjadi huruf kecil, *tokenizing* mengurai kata, *filtering* menyisikan kata-kata *stopword*, dan *stemming* menghasilkan kata dasar.
3. Kata kunci, hasil dari pra-pengolahan menghasilkan kata kunci untuk diolah ke dalam pencarian perhitungan jarak.
4. Pra-pengolahan, pada tahap *document retrieval* ini dokumen akan melewati proses *case folding* dengan mengubah karakter menjadi huruf kecil, *tokenizing* mengurai kata, *filtering* menyisikan kata-kata *stopword*, *stemming* menghasilkan kata dasar, serta pengambilan kalimat-kalimat definisi yang mengandung kata kunci hingga menghasilkan dokumen terbaik.
5. *Search*, tahap ini melakukan proses pencarian dokumen terbaik. Dokumen terbaik tersebut nantinya akan berisikan hasil penggalan jawaban terbaik yang dikeluarkan oleh sistem dari perhitungan TF-IDF dengan kata kunci dan hasil perhitungan jawaban terbaik.
6. Menghitung jarak teks jawaban dengan algoritma metode pendekatan jarak, yaitu *Manhattan Distance*, *Euclidean Distance*, dan *Cosine Similarity*. Jarak teks dihitung untuk mendapatkan jawaban terbaik dari pertanyaan.
7. Memilih dan menampilkan jawaban dari kandidat yang sering muncul.

Contoh perhitungan pada Tabel III-3 adalah contoh pertanyaan dan jawaban untuk perhitungan nilai TF-IDF pada Tabel III-4.

Tabel III-3. Contoh Pertanyaan dan Jawaban

Pertanyaan	Apa yang dimaksud dengan nuklir?
Pertanyaan Hasil Pra-pengolahan	Energi nuklir
Jawaban	Energi nuklir, yaitu energi yang terdapat pada inti atom?
Jawaban Hasil Pra-pengolahan	Energi nuklir inti atom

Perhitungan pada TF-IDF dimulai dengan menghitung nilai TF serta nilai IDF pada setiap pertanyaan dan jawaban yang telah di pra-pengolahan, Setelah mendapatkan masing-masing nilai tersebut nilai TF akan dikalikan dengan nilai IDF dengan rumus sebagai berikut:

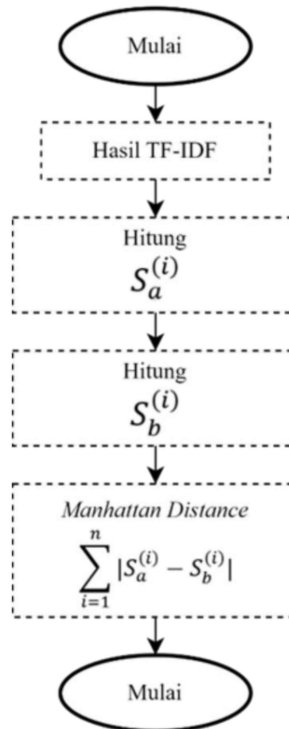
$$tf_{(t,d)} = \frac{\text{number of times } t \text{ appears in } d}{\text{total number of terms in } d}$$

$$idf_{(t)} = \log \frac{N}{N_{(t)}}$$

$$tf - idf_{(t,d)} = tf_{(t,d)} \times idf_{(t)}$$

Tabel III-4. Perhitungan TF-IDF

Kata	TF		IDF	TF-IDF	
	P	J		P	J
energi	1/8	0/16	$\text{Log}(2/2) = 0$	0	0
nuklir	1/8	0/16	$\text{Log}(2/2) = 0$	0	0
inti	1/8	0/16	$\text{Log}(2/1) = 0,301$	0	0,0752
atom	2/8	2/16	$\text{Log}(2/1) = 0,301$	0	0,0752

a. Perhitungan *Manhattan Distance*

Gambar III-3. Flowchart perhitungan *Manhattan Distance*

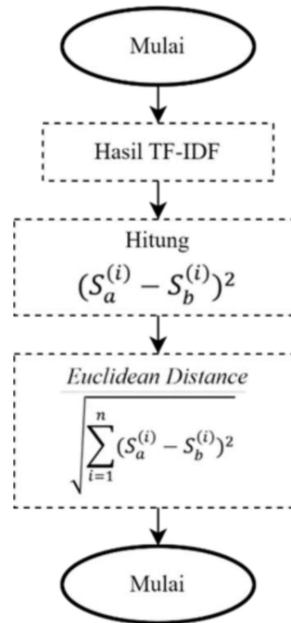
Pada Gambar III-3, pengujian akan melakukan perhitungan TF-IDF dan setelahnya menghitung *Manhattan Distance*. Perhitungan jarak dapat dihitung sebagai berikut:

$$d(S_a, S_b) = \sum_{i=1}^n |S_a^{(i)} - S_b^{(i)}|$$

$$d(S_a, S_b) = |(0) - (0)| + |(0) - (0)| + |(0) - (0,0752)| + |(0) - (0, 0752)|$$

$$d(S_a, S_b) = |-0,15051|$$

$$d(S_a, S_b) = 0,15051$$

b. Perhitungan *Euclidean Distance*

Gambar III-4. Flowchart perhitungan *Euclidean Distance*

Pada Gambar III-4, pengujian akan melakukan perhitungan TF-IDF dan setelahnya menghitung *Euclidean Distance*. Perhitungan jarak dapat dihitung sebagai berikut:

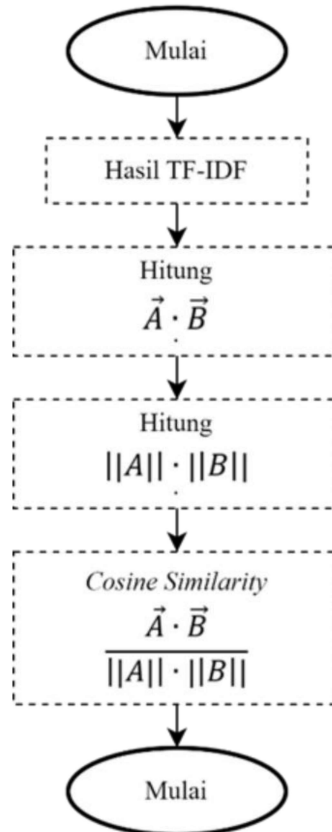
$$d(S_a, S_b) = \sqrt{\sum_{i=1}^n (S_a^{(i)} - S_b^{(i)})^2}$$

$$d(S_a, S_b) = \sqrt{(0 - 0)^2 + (0 - 0)^2 + (0 - 0,0752)^2 + (0 - 0,0752)^2}$$

$$d(S_a, S_b) = \sqrt{0 + 0 + 0,00565 + 0,00565}$$

$$d(S_a, S_b) = \sqrt{0,01131}$$

$$d(S_a, S_b) = 0,10634$$

c. Perhitungan *Cosine Similarity*

Gambar III-5. Flowchart perhitungan *Cosine Similarity*

Pada Gambar III-5, pengujian akan melakukan perhitungan TF-IDF dan setelahnya menghitung *Cosine Similarity*.

$$sim = \frac{\vec{A} \cdot \vec{B}}{\|A\| \cdot \|B\|}$$

$$sim = \frac{(0 \times 0) + (0 \times 0) + (0 \times 0,0753) + (0 \times 0,0753)}{\sqrt{0^2 + 0^2 + 0^2 + 0^2} \times \sqrt{0^2 + 0^2 + 0,0753^2 + 0,0753^2}}$$

$$sim = \frac{0 + 0 + 0 + 0}{\sqrt{0} \times \sqrt{0,01134}}$$

$$sim = \frac{0}{0 \times 0,10649}$$

$$sim = 0$$

Dalam membangun Sistem Tanya Jawab untuk pertanyaan apa menggunakan metode pendekatan jarak ini, terdapat tahapan menetapkan kriteria pengujian sebagai berikut.

1. Keluaran sistem dinyatakan tepat dengan fakta jawaban dari pertanyaan.
2. Jawaban memiliki pola sesuai dengan yang telah ditetapkan.
3. Jawaban bertipe pertanyaan *non-factoid*.

3.3.3 Menentukan Alat Bantu Penelitian

Alat bantu yang digunakan pada penelitian Sistem Tanya Jawab untuk pertanyaan apa ialah sebagai berikut:

1. Perangkat Keras (*Hardware*)

Prosesor : AMD Ryzen 7 4800H with Radeon Graphics 2.90 GHz

RAM : 16.0 GB

SSD : 512 GB

2. Perangkat Lunak (*Software*)

Bahasa Pemograman : Python dan Javascript

Framework : Flask

Pustaka Eksternal : flask, glob, numpy, pandas, regex, dan NLTK

3.3.4 Melakukan Pengujian

Pengujian dalam penelitian Sistem Tanya Jawab untuk pertanyaan apa menggunakan metode pendekatan jarak dilakukan berdasarkan kriteria pengujian pada tahap membangun sistem. Data pengujian akan disajikan dalam bentuk tabel *confusion matrix*. Tabel III-5 merupakan format data yang digunakan dalam evaluasi hasil penelitian.

Tabel III-5. Pengujian Pertanyaan

No	Pertanyaan	Fakta Jawaban	Output	Ket			
				TP	FP	TN	FN
1.							
2.							

Tabel pengujian terbagi atas kolom pertanyaan “apa”, kolom fakta jawaban berisikan jawaban benar yang seharusnya menjawab pertanyaan tersebut, dan kolom *output* berisikan keluaran jawaban yang diberikan oleh sistem. Kolom keterangan berisikan hasil dari TP, TN, FP, dan FN yang merupakan nilai representasi hasil proses klasifikasi yang dihasilkan oleh sistem QA. Nilai tersebut akan digunakan untuk menghitung nilai *accuracy*, *precision*, dan *recall*.

Keterangan komponen *confusion matrix* dalam pengujian ini dapat dijabarkan sebagai berikut:

1. True Positive : Merupakan jawaban benar pada kategori jawaban terbaik algoritma pendekatan jarak yang diambil dari dataset.

2. False Positive : Merupakan jawaban salah pada kategori jawaban terbaik algoritma pendekatan jarak yang diambil dari dataset.
3. True Negative : Merupakan jawaban benar pada kategori jawaban terbaik algoritma pendekatan jarak bukan bersumber dari dataset.
4. False Negative : Merupakan jawaban salah pada kategori jawaban terbaik algoritma pendekatan jarak bukan bersumber dari dataset.

Pengujian pertanyaan akan dibagi menjadi tiga masing-masing untuk membandingkan hasil penilaian yang didapatkan dari perbandingan algoritma metode pendekatan jarak, yaitu *Manhattan Distance*, *Euclidean Distance*, dan *Cosine Similarity*.

3.3.5 Menganalisis Hasil Pengujian dan Menarik Kesimpulan

Pengujian ini akan dianalisis berdasarkan nilai perhitungan jarak terdekat yang dihasilkan sistem. Pengujian ini akan dimulai dengan menghitung nilai TF-IDF pada pertanyaan, dokumen, dan jawaban. Hasil perhitungan nilai TF akan dikalikan dengan nilai IDF, dari perhitungan ini akan menghasilkan nilai vektor yang akan digunakan untuk perhitungan algoritma pendekatan jarak.

Nilai vektor TF-IDF ini akan menjadi nilai untuk menghitung algoritma *Manhattan Distance*, *Euclidean Distance*, dan *Cosine Similarity*. Hasil perhitungan pada setiap algoritmanya akan di keluarkan oleh sistem, nilai yang di keluarkan adalah satu nilai terbesar untuk *Cosine Similarity* serta nilai terkecil untuk *Manhattan Distance* dan *Euclidean Distance*, yaitu nilai yang dianggap nilai terdekat.

Perhitungan algoritma ini digunakan dalam dari perhitungan setiap pertanyaan dengan satu per satu jawaban didalam dataset ini, sistem akan mengambil nilai hasil perhitungan yang terbaik. Hasil perhitungan dengan nilai terbesar untuk *Cosine Similarity* serta nilai terkecil untuk *Manhattan Distance* dan *Euclidean Distance* tersebut merupakan jawaban terbaik yang ditampilkan oleh sistem.

Keluaran sistem kemudian dibandingkan dengan jawaban sebenarnya dari pasangan pertanyaan dan jawaban pada dataset dengan menggunakan klasifikasi *confusion matrix* menentukan nilai TP, TN, FP, dan FN, serta diberi bobot setiap keluaran jawaban yang dihasilkan oleh sistem. Bobot tersebut akan digunakan dalam menghitung nilai *accuracy*, nilai *precision*, dan nilai *recall* sebagai kesimpulan akhir penelitian.

3.3.6 Membuat Laporan Penelitian

Tahap membuat laporan penelitian merupakan tahap akhir yang dilakukan untuk memberikan informasi seputar penelitian yang dilakukan dengan menggambarkan hasil dari penelitian dalam bentuk dokumen atau laporan penelitian. Laporan penelitian berisikan deskripsi tujuan penelitian yang dilakukan, metodologi, hasil, serta kesimpulan dan saran. Laporan penelitian disusun berdasarkan metode penulisan dan sistematika yang telah ditentukan. Tujuan laporan ini adalah membuat pembaca dan penulis benar-benar memahami bahan penelitian yang dilakukan.

3.4 Manajemen Proyek Penelitian

Pada subbab manajemen proyek penelitian ini membahas perencanaan kegiatan penelitian dari tahap identifikasi masalah hingga tahap menarik kesimpulan dari penelitian yang telah dilakukan. Proses perencanaan kegiatan penelitian tersebut dapat dibuat dalam bentuk *Gantt Chart* yang dapat dilihat pada Tabel III-3.(pada lampiran 1)

3.5 Kesimpulan

Pada bab ini telah membahas mengenai tahapan penelitian yang dilakukan, terdiri atas mengumpulkan data, menentukan kriteria dan format data yang diperlukan dalam proses pengujian serta analisis hasil pengujian

BAB IV

PENGEMBANGAN PERANGKAT LUNAK

4.1 Pendahuluan

Pada bab ini akan membahas mengenai pengembangan perangkat lunak yang digunakan pada penelitian. Metode yang digunakan adalah *waterfall*, dengan fase-fase sebagai berikut.

4.2 *Waterfall*

Dalam penelitian ini membangun perangkat lunak dengan menggunakan metode *waterfall* yang terdiri dari empat fase, yaitu *requirement*, *design*, *implementation*, *verification* dan *maintenance*.

4.2.1 Requirement

Tahapan *requirement* ini terdapat pembagian kebutuhan sistem yang terbagi menjadi dua kategori, yaitu kebutuhan fungsional dan kebutuhan non-fungsional sistem. Kebutuhan fungsional menjelaskan proses yang dilakukan oleh sistem dan mengidentifikasi fasilitas yang diperlukan, sebagaimana yang terlihat pada Tabel IV-1. Sementara itu, kebutuhan non-fungsional lebih menekankan pada atribut-atribut perilaku yang diinginkan untuk mencapai tujuan sistem yang diinginkan, sebagaimana yang terlihat pada Tabel IV-2.

Tabel IV-1. Tabel Kebutuhan Fungsional

No.	Kebutuhan Fungsional
1.	Sistem dapat melakukan input pertanyaan.
2.	Sistem dapat melakukan pra-pengolahan pada pertanyaan.
3.	Sistem dapat melakukan pra-pengolahan pada dataset dalam format .txt.
4.	Sistem melakukan perhitungan vektor TF-IDF
5.	Sistem dapat melakukan proses pencarian jawaban dengan hasil perhitungan terdekat dari 3 metode.
6.	Sistem dapat menampilkan 3 jawaban terbaik dari setiap metode.

Tabel IV-2. Tabel Kebutuhan Non-Fungsional

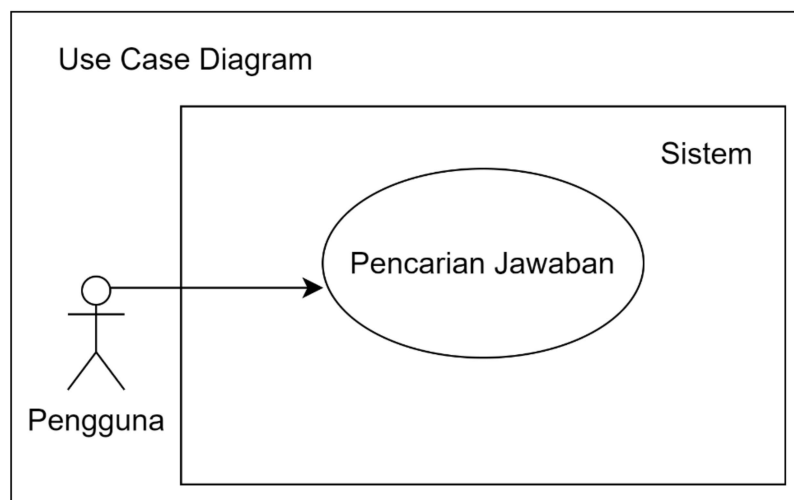
No.	Kebutuhan Non-Fungsional
1.	Sistem memiliki tampilan antar muka yang mudah dipahami dan digunakan oleh pengguna.

4.2.2 *Design*

Tahap ini berisi perancangan sistem yang dibuat dalam bentuk diagram, seperti *use case diagram*, *activity diagram* dan *sequence diagram*.

4.2.2.1 *Use Case Diagram*

Use Case Diagram merupakan representasi visual yang memberikan gambaran fungsional bagaimana suatu sistem bekerja dengan tujuan agar pengguna dapat memahami atau mengerti cara kerja suatu sistem. *Use Case* pada penelitian ini dapat dilihat pada Gambar IV-1.



Gambar IV-1. Diagram *Use Case*

a. Definisi *Actor*

Aktor dalam penelitian ini adalah pengguna yang akan berinteraksi dengan sistem yang dibuat. Pengguna dapat menggunakan semua fitur yang ada pada sistem.

Tabel IV-3 merupakan table definisi *actor*.

Tabel IV-3. Definisi Actor

No.	<i>Actor</i>	Definisi
1.	Pengguna	Pengguna adalah pengguna yang akan berinteraksi secara langsung dengan sistem dan menggunakan fitur pada sistem.

b. Definisi *Use Case*

Deskripsi *Use Case* pada sistem ini dapat dilihat pada Tabel IV-4.

Tabel IV-4. Definisi *Use Case*

No.	<i>Use Case</i>	Definisi
1.	Pencarian Jawaban	Melakukan pra-pengolahan dan dilanjutkan melakukan perhitungan vektor TF-IDF serta jarak terdekat untuk menghasilkan jawaban terbaik yang akan ditampilkan.

c. Skenario *Use Case*

Skenario *Use Case* menggambarkan aktor yang melakukan interaksi terhadap sistem, dan menggambarkan respon yang diberikan sistem terhadap prosedur yang dilakukan oleh aktor.

Skenario pertama pada Tabel IV-5 merupakan *use case* pengguna dalam melakukan pencarian pertanyaan, pertanyaan tersebut akan diolah sistem agar dapat mengeluarkan jawaban yang paling tepat dalam setiap algoritma pendekatan jarak yang digunakan.

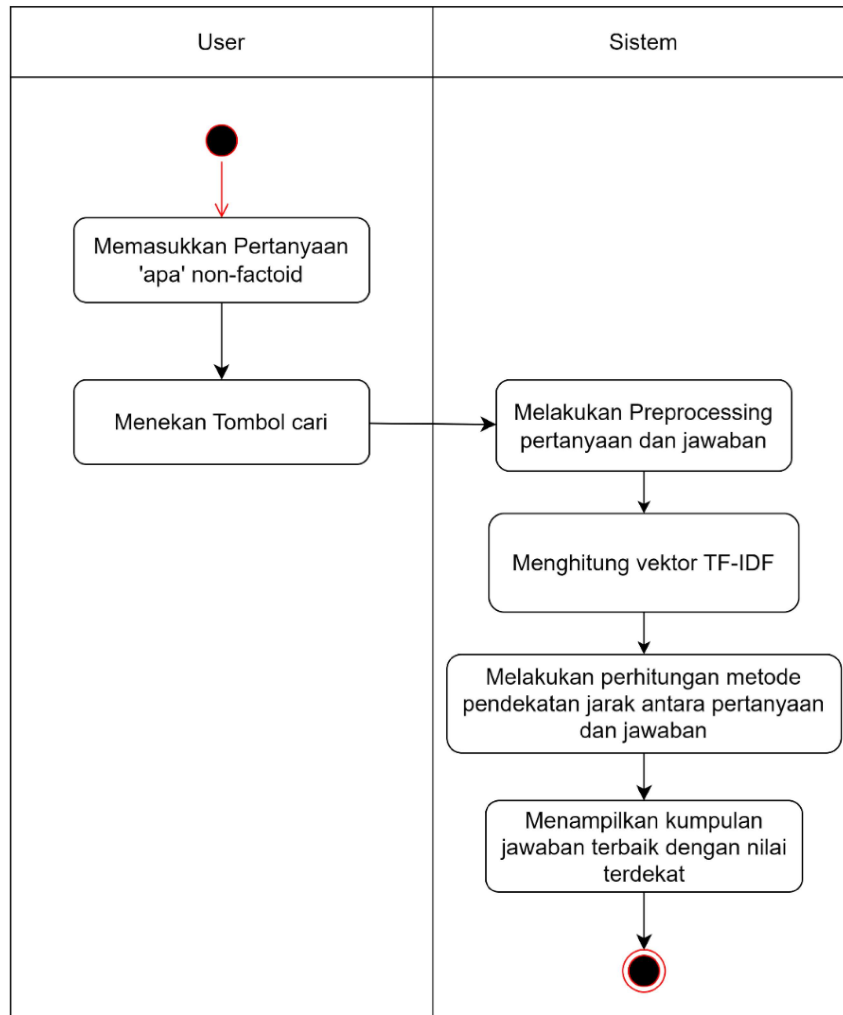
Tabel IV-5. Skenario *Use Case* Melakukan Pencarian Jawaban

Identifikasi	
Nomor	001
Nama	Melakukan Pencarian Jawaban
Actor	Pengguna
Tujuan	Sistem dapat menghasilkan jawaban terbaik dengan algoritma pendekatan jarak

Deskripsi	Melakukan proses pencarian jawaban berdasarkan pertanyaan masukan
Kondisi Awal	Tampilan halaman awal berisi kolom pencarian dan tombol pencarian
Skenario Normal	
Aksi Aktor	Reaksi Sistem / Perangkat lunak
5. Memasukkan pertanyaan pada kolom pertanyaan. 6. Menekan tombol pencarian	
	<ol style="list-style-type: none"> 1. Melakukan pra-pengolahan pertanyaan masukan pengguna dan jawaban dari <i>dataset</i> 2. Menghitung vektor TF-IDF 3. Melakukan perhitungan setiap algoritma pendekatan jarak 4. Mengambil jawaban dengan nilai terdekat antara pertanyaan dan jawaban pada setiap metode 5. Mengumpulkan dan menampilkan jawaban-jawaban terbaik
Kodisi Akhir	Sistem menampilkan jawaban terbaik pada setiap algoritma pendekatan jarak

4.2.2.2 Activity Diagram

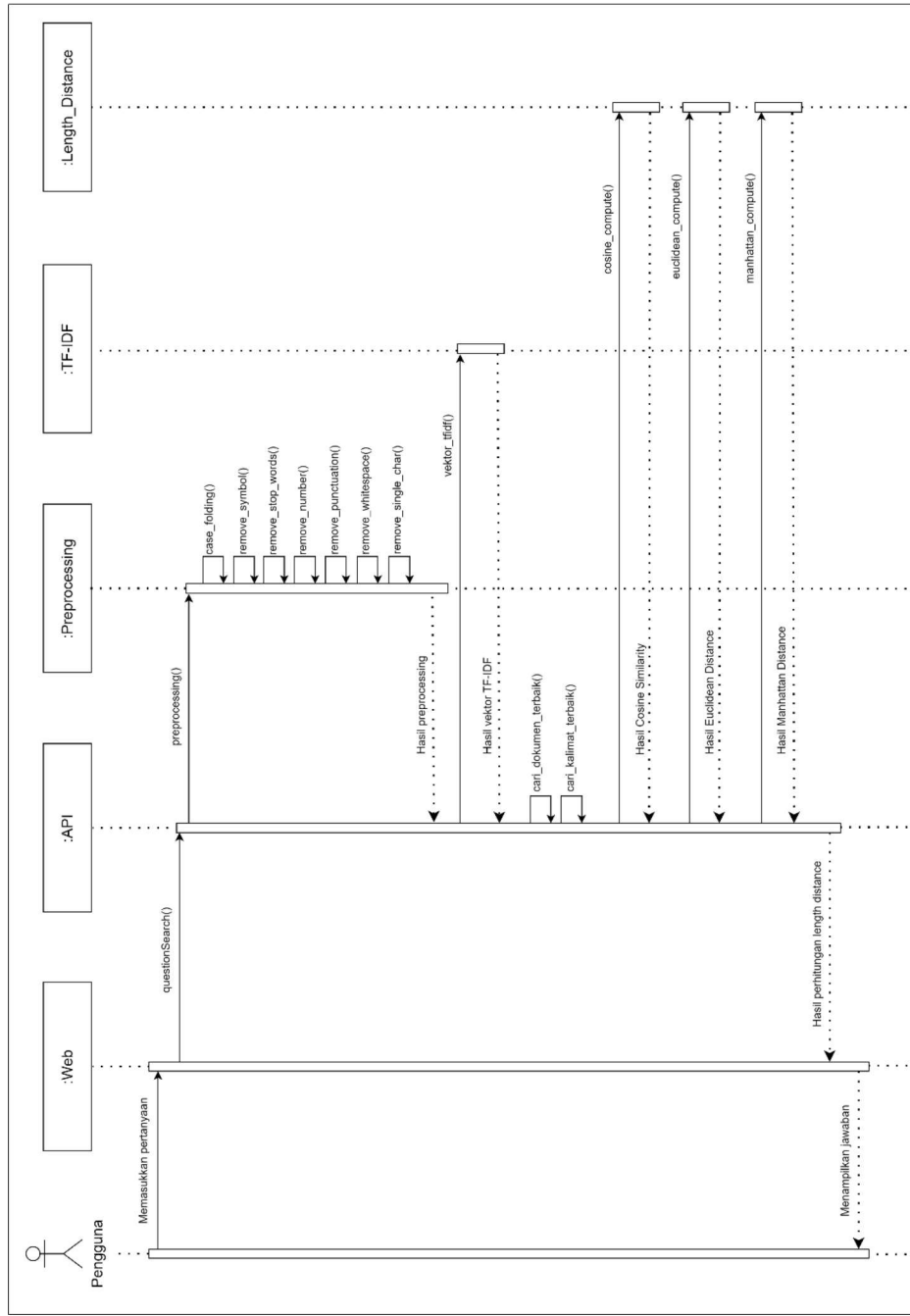
Activity Diagram yang dipresentasikan pada Gambar IV-2 merupakan diagram yang digunakan dalam menggambarkan alur kerja sistem yang terjadi pada sistem yang dibangun. Diagram ini menjelaskan aktivitas serta langkah-langkah interaksi *actor* pada sebuah sistem.



Gambar IV-2. *Activity Diagram* Melakukan Pencarian Jawaban

4.2.2.3 *Sequence Diagram*

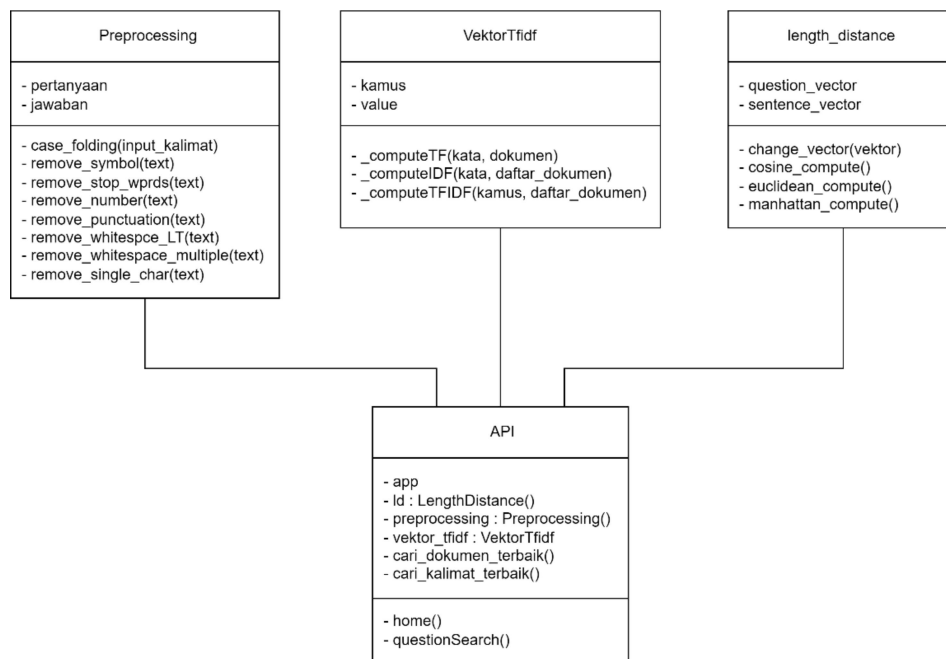
Sequence Diagram pada Gambar IV-3 digunakan untuk menggambarkan dan menampilkan interaksi antar objek dalam suatu sistem secara terperinci dalam urutan waktu. Diagram ini memberikan gambaran visual tentang aliran kontrol dan komunikasi antara objek-objek dalam sistem.



Gambar IV-3. Sequence Diagram Proses Pencarian Jawaban dengan Perhitungan Jarak Terdekat

4.2.2.4 Class Diagram

Class Diagram pada Gambar IV-4 memberikan gambaran *structural* tentang kelas-kelas, atribut-atribut, dan hubungan antar kelas pada sistem.



Gambar IV-4. *Class Diagram*

4.2.2.5 Perancangan Antar Muka

Perancangan antar muka perangkat lunak ini menampilkan sebuah halaman yang berisikan sebuah kolom untuk mengisikan pertanyaan yang disertai tombol untuk mencari pertanyaan tersebut.

Pada Gambar IV-5 dapat dilihat tampilan sederhana untuk penelitian ini yang nantinya jika tombol “cari” ditekan akan mengeluarkan jawaban terbaik dari masing masing metode.



Gambar IV-5. *Wireframe*

4.2.3 Implementation

Tahap implementasi ini merupakan tahap pemrograman, membuat sebuah desain ke dalam bentuk kodingan hingga menghasilkan bentuk yang sama dengan perencanaan. Pada Tabel IV-6 dan Gambar IV-6 dapat dilihat hasil implementasi terhadap perangkat lunak ini.

Tabel IV-6. Tabel Implementasi Objek

No.	Nama Objek	Implementasi dalam Bentuk <i>File</i>	Keterangan
1.	API	app.py	File “app.py” ini berisikan API yang untuk menampilkan halaman <i>website</i> serta fungsi <code>questionSearch()</code> yang merupakan proses pengambilan pertanyaan dan jawaban untuk menghitung pra-pengolahan, TF-IDF, serta algoritma pendekatan jarak untuk

			menghasilkan tampilan jawaban terbaik di web.
2.	Pra-pengolahan	pra-pengolahan.py	<i>File “preprocessing.py”</i> berisikan pra-pengolahan yang akan digunakan pada pertanyaan dan jawaban yang nantinya akan digunakan dalam perhitungan TF-IDF. Pra-pengolahan ini mengolah data mentah sehingga menghasilkan data bersih untuk pengujian.
3.	TF-IDF	vektor_tfidf.py	<i>File “vektor_tfidf.py”</i> berisikan perhitungan TF-IDF untuk memperoleh nilai vektor yang akan digunakan dalam perhitungan algoritma pendekatan jarak
4.	Length_Distance	length_distance.py	<i>File “length_distance.py”</i> ini berisikan perhitungan algoritma pendekatan jarak, yaitu <i>Manhattan Distance</i> , <i>Euclidean Distance</i> , dan <i>Cosine Similarity</i> .

QUESTION ANSWERING SYSTEM

Pertanyaan 'Apa' Non-Factoid

CARI

Cosine Similarity

Lorem ipsum, atau ringkasnya lipsum, adalah teks standar yang ditempatkan untuk mendemostrasikan elemen grafis atau presentasi visual seperti font,

Euclidean Distance

Lorem ipsum, atau ringkasnya lipsum, adalah teks standar yang ditempatkan untuk mendemostrasikan elemen grafis atau presentasi visual seperti font,

Manhattan Distance

Lorem ipsum, atau ringkasnya lipsum, adalah teks standar yang ditempatkan untuk mendemostrasikan elemen grafis atau presentasi visual seperti font,

Gambar IV-6. Implementasi antar muka

4.2.4 Verification

Verifikasi pada penelitian ini dilakukan dengan menguji fungsionalitas sistem dengan menggunakan metode *black box testing* agar lebih fokus ke dalam perspektif pengguna akhir dalam memastikan sistem berperilaku sesuai dengan kebutuhan dan harapan. Pengujian sistem dilakukan pada fungsionalitas tombol. Proses pengujian ini dapat dilihat pada Tabel IV-7.

Tabel IV-7. Tabel Pengujian Sistem

No.	Prosedur Pengujian	Deskripsi	Output yang diharapkan	Hasil Output dari Sistem	Kesimpulan
1.	Melakukan klik tombol "Cari"	Melakukan klik pada tombol "Cari" yang digunakan untuk memroses pertanyaan	Menampilkan hasil jawaban terbaik dari pertanyaan	Menampilkan jawaban dengan nilai pendekatan terkecil sesuai perhitungan di setiap metodenya	Berhasil

4.2.5 Maintenance

Dalam pembuatan sistem ini, tahap perawatan tidak diperlukan dalam implementasi karena sistem dirancang untuk memenuhi kebutuhan pengguna dalam mempermudah penelitian.

4.3 Kesimpulan

Pada bab ini menjelaskan mengenai penjabaran metodologi *waterfall* yang digunakan dalam pengembangan perangkat lunak oleh peneliti, yakni Sistem Tanya Jawab untuk Pertanyaan Apa Menggunakan Metode Pendekatan Jarak.

BAB V

HASIL DAN ANALISIS PENELITIAN

5.1 Pendahuluan

Pada bab ini menjelaskan mengenai hasil dan analisis implementasi pengujian dataset dan konfigurasi percobaan dari penelitian Sistem Tanya Jawab untuk Pertanyaan Apa Menggunakan Metode Pendekatan Jarak.

5.2 Data Hasil Percobaan Penelitian

Percobaan penelitian ini dilakukan dengan menguji dataset sebanyak 110 pasang pertanyaan dan jawaban serta 61 dokumen yang telah melewati pemrosesan data hingga mendapatkan kata kunci sebelum dilanjutkan ke proses perhitungan metode. Menggunakan pasangan pertanyaan dan jawaban yang tepat tersebut, penelitian ini dapat menghitung nilai akurasi dari keluaran jawaban yang dihasilkan oleh sistem.

5.2.1 Konfigurasi Percobaan

Data yang digunakan dalam pengujian ini berjumlah 110 pasang pertanyaan ‘apa’ *non-factoid* beserta jawabannya tepatnya dalam bentuk format .txt. Jawaban akan diperoleh dari sepenggal kalimat dengan kedekatan terbaik pada hasil perhitungan dokumen terbaik. Data pertanyaan dan dokumen tersebut akan melalui tahap-tahap pra-pengolahan dalam memperoleh kata kunci, yaitu mencakup tahap *case folding*, *tokenization*, *filtering*, dan *stemming*. Selanjutnya, setiap kata kunci yang dihasilkan dari proses sebelumnya ini akan digunakan dalam

perhitungan nilai TF-IDF sebagai penilaian komprehensif tentang pentingnya sebuah kata dalam dokumen atau teks tersebut. Perhitungan jarak teks akan dihitung dengan algoritma pendekatan jarak, yaitu *Manhattan Distance*, *Euclidean Distance*, dan *Cosine Similarity* dari nilai TF-IDF yang didapatkan dalam perhitungan sebelumnya. Keluaran nilai terbesar untuk *Cosine Similarity* serta nilai terkecil untuk *Manhattan Distance* dan *Euclidean Distance* dari setiap metode tersebut dapat dijadikan kandidat jawaban terbaik karena jaraknya yang dekat.

Kandidat jawaban terbaik dari setiap metode tersebut akan dibandingkan dengan jawaban tepat dari pasangan pertanyaan dan jawaban pada dataset. Hasil bobot dari perbandingan tersebut akan memperoleh nilai TP, TN, FP, dan FN yang digunakan untuk menghitung nilai *Accuracy*, *Recall*, dan *Precision*.

Nilai *confusion matrix* yang dihasilkan dari pengujian dalam penelitian ini hanya *True Positive* (TP) dan *False Positive* (FP). Nilai *True Positive* (TP) dan *False Positive* (FP) memiliki bobot karena pada penelitian ini jawaban yang dikeluarkan sistem akan selalu terdapat pada dataset, akan tetapi belum tentu memiliki keluaran yang selalu tepat. Keluaran yang dihasilkan sistem tidak mungkin keluaran *negative* yaitu jawaban yang tidak terdapat pada dataset, sehingga nilai *True Negative* (TN) dan *False Negative* (FN) tidak memiliki bobot nilai.

Pada penelitian ini dilakukan dengan 4 skenario konfigurasi percobaan.

Adapun scenario pada konfigurasi percobaan pada penelitian ini sebagai berikut:

1. Pada skenario konfigurasi percobaan 1, data pertanyaan dan data dokumen yang telah menghasilkan kata-kata penting dari tahap pra-pengolahan, akan langsung dihitung nilai TF-IDF dan algoritma-algoritma pendekatan jaraknya untuk menghasilkan dokumen terbaik. Dokumen terbaik yang dihasilkan akan di pra-pengolahan pada setiap kalimatnya untuk menghasilkan kata-kata penting yang akan dihitung nilai TF-IDF dan algoritma-algoritma pendekatan jaraknya dalam memperoleh jawaban terbaik. Jawaban terbaik akan dibandingkan dengan jawaban sebenarnya untuk menentukan *True Positive* (TP) dan *False Positive* (FP).
2. Pada skenario konfigurasi percobaan 2, data pertanyaan dan data dokumen dipra-pengolahan dan akan dihitung nilai TF-IDF serta algoritma-algoritma pendekatan jaraknya. Setelah menghasilkan dokumen terbaik tersebut, dokumen akan di *filter* untuk mengambil kalimat-kalimat definisi yang mengandung kata kunci ‘adalah’, ‘yaitu’, ‘ialah’, ‘merupakan’, ‘diartikan’, ‘artinya’, ‘disebut’, ‘dikenal’, ‘dinamakan’, ‘mendefinisikan’. Kalimat yang sudah di *filter* ini akan masuk ke tahap pra-pengolahan untuk memperoleh kata-kata penting dan melakukan perhitungan TF-IDF dan algoritma-algoritma pendekatan jarak untuk menghasilkan jawaban terbaik. Jawaban terbaik akan dibandingkan dengan jawaban sebenarnya untuk menentukan *True Positive* (TP) dan *False Positive* (FP).

3. Pada skenario konfigurasi percobaan 3, akan dilakukan sama dengan skenario konfigurasi percobaan 2. Jawaban terbaik yang diperoleh akan diambil kata-kata penting didalamnya dan dibandingkan dengan kata-kata penting didalam jawaban sebenarnya. Jika kata-kata penting pada jawaban terbaik mengandung 50% kata-kata penting pada jawaban sebenarnya maka akan dianggap *True Positive* (TP).
4. Pada skenario konfigurasi percobaan 4, akan dilakukan sama dengan skenario konfigurasi percobaan 3, hanya saja kata-kata penting pada jawaban terbaik harus mengandung 40% kata-kata penting pada jawaban sebenarnya untuk dianggap *True Positive* (TP).

5.2.2 Data Hasil Pengujian *Cosine Similarity*

Setelah dilakukan pengujian menggunakan metode *Cosine Similarity* dengan kata kunci yang diperoleh pada tahap pra-pengolahan dan perhitungan TF-IDF sebelumnya, dokumen terbaik akan dihasilkan oleh keluaran sistem berdasarkan nilai terbesar yang dianggap nilai terbaik seperti pada Tabel V-1.

Tabel V-1. Dokumen Terbaik *Cosine Similarity*

Pertanyaan	Kata Kunci Pertanyaan	Dokumen Terbaik (.txt)
Apa yang dimaksud dengan Albuminuria?	albuminuria	43
Apa yang dimaksud dengan Defekasi?	defekasi	43
Apakah yang dimaksud dengan Protoplasma ?	protoplasma	36
Apakah yang dimaksud dengan trinomial?	trinomial	26
Apakah yang dimaksud dengan protista?	protista	29

Tabel V-2. Pengujian *Cosine Similarity*

Pengujian	Confusion Matrix				Accuracy	Recall	Precision
	TP	TN	FP	FN			
Skenario 1	23	0	87	0	0.209090909090908	1	0.209090909090908
Skenario 2	23	0	87	0	0.209090909090908	1	0.209090909090908
Skenario 3	35	0	75	0	0.31818181818182	1	0.31818181818182
Skenario 4	41	0	69	0	0.372727272727274	1	0.372727272727274

Dari hasil Tabel V-2 dapat diketahui metode *Cosine Similarity* memperoleh hasil terbaik pada scenario konfigurasi percobaan 4 dengan hasil *accuracy* dan *precision* sebesar 37.27 %, serta nilai *recall* 100%. Bobot *True Positive* terbaik yang dihasilkan pada

perhitungan dengan algoritma *Cosine Similarity* ini hanya sebesar 41, sedangkan bobot *False Negative* sebesar 69. Seperti yang dijabarkan pada Tabel V-3 adalah hasil pengujian algoritma *Cosine Similarity*.

Tabel V-3. Nilai Output *Cosine Similarity*

No.	Pertanyaan	Jawaban	Keluaran Sistem	Ket.
1.	Apa yang dimaksud dengan Albuminuria?	Albuminuria, merupakan gangguan yang terjadi pada ginjal sehingga urine mengandung protein. Di dalam urine normal, sebenarnya tidak mengandung senyawa protein, asam amino, ataupun glukosa. Oleh karena itu, gangguan ini menunjukkan bila alat filtrasi pada ginjal telah rusak.	Albuminuria, merupakan gangguan yang terjadi pada ginjal sehingga urine mengandung protein	TP (0.497252340479922)
2.	Apa yang dimaksud dengan Defekasi?	Defekasi merupakan proses pengeluaran sisa-sisa pencernaan makanan yang disebut feces melalui anus.	Defekasi merupakan proses pengeluaran sisa-sisa pencernaan makanan yang disebut feces melalui anus	TP (0.43347341920785254)

3.	Apakah yang dimaksud dengan Protoplasma ?	Protoplasma adalah sejenis substansi kompleks seperti agar-agar yang tidak habis digunakan saat aktivitas kimiawi dalam menjaga kelangsungan hidup sel.	Protoplasma merupakan bagian sel yang berisi cairan menyerupai agar-agar	FP (0.5672845056144695)
4.	Apakah yang dimaksud dengan trinomial?	Sistem penamaan yang terdiri atas tiga suku kata disebut Trinomial, contohnya, Passer domesticus domesticus (burung gereja) dan Felis maniculata domesticus (kucing jinak).	Klasifikasi adalah pengelompokan makhluk hidup dalam takson melalui pencarian keseragaman atau persamaan dalam keanekaragaman	FP (0.0)
5.	Apakah yang dimaksud dengan protista?	Protista berasal dari bahasa Yunani protos yang berarti 'pertama'. Jadi, Protista merupakan eukariotik pertama hasil evolusi prokariotik.	Anggota sporozoa yang paling dikenal adalah Plasmodium	FP (0.5440146428047532)

5.2.3 Data Hasil Pengujian *Euclidean Distance*

Setelah dilakukan pengujian menggunakan metode *Euclidean Distance* dengan kata kunci yang diperoleh pada pra-pengolahan dan perhitungan TF-IDF sebelumnya, akan menampilkan dokumen terbaik dari hasil perhitungan dengan nilai terkecil dan memperoleh hasil seperti pada Tabel V-4.

Tabel V-4. Dokumen Terbaik *Euclidean Distance*

Pertanyaan	Kata Kunci Pertanyaan	Dokumen Terbaik (.txt)
Apa yang dimaksud dengan Albuminuria?	albuminuria	43
Apa yang dimaksud dengan Defekasi?	defekasi	43
Apakah yang dimaksud dengan Protoplasma ?	protoplasma	36
Apakah yang dimaksud dengan trinomial?	trinomial	26
Apakah yang dimaksud dengan protista?	protista	29

Tabel V-5. Pengujian *Euclidean Distance*

Pengujian	Confusion Matrix				Accuracy	Recall	Precision
	TP	TN	FP	FN			
Skenario 1	23	0	87	0	0.209090909090908	1	0.209090909090908
Skenario 2	21	0	89	0	0.190909090909092	1	0.190909090909092
Skenario 3	34	0	76	0	0.3090909090909091	1	0.3090909090909091
Skenario 4	40	0	70	0	0.363636363636365	1	0.363636363636365

Dari hasil Tabel V-5 dapat diketahui metode *Euclidean Distance* memperoleh hasil terbaik pada skenario konfigurasi percobaan 4 dengan hasil *accuracy* dan *precision* sebesar 36.36%, serta nilai *recall* 100%. Bobot *True Positive* terbaik yang dihasilkan pada perhitungan dengan algoritma *Euclidean Distance* ini hanya sebesar 40, sedangkan bobot *False Negative* sebesar 70. Seperti yang dijabarkan pada Tabel V-6 adalah hasil pengujian algoritma *Euclidean Distance*.

Tabel V-6. Nilai Output Euclidean Distance

No.	Pertanyaan	Jawaban	Keluaran Sistem	Ket.
1.	Apa yang dimaksud dengan Albuminuria?	Albuminuria, merupakan gangguan yang terjadi pada ginjal sehingga urine mengandung protein. Di dalam urine normal, sebenarnya tidak mengandung senyawa protein, asam amino, ataupun glukosa. Oleh karena itu, gangguan ini menunjukkan bila alat filtrasi pada ginjal telah rusak.	Albuminuria, merupakan gangguan yang terjadi pada ginjal sehingga urine mengandung protein	TP (1.4839991486743753)
2.	Apa yang dimaksud dengan Defekasi?	Defekasi merupakan proses pengeluaran sisa pencernaan makanan yang disebut feses melalui anus.	Defekasi merupakan proses pengeluaran sisa-sisa pencernaan makanan yang disebut feses melalui anus	TP (1.567606771018028)
3.	Apakah yang dimaksud dengan Protoplasma ?	Protoplasma adalah sejenis substansi kompleks seperti agar-agar yang tidak habis digunakan saat aktivitas kimiawi dalam menjaga kelangsungan hidup sel.	Protoplasma merupakan bagian sel yang berisi cairan menyerupai agar-agar	FP (0.8527825607625409)

<p>4.</p>	<p>Apakah yang dimaksud dengan trinomial?</p>	<p>Sistem penamaan yang terdiri atas tiga suku kata disebut Trinomial, contohnya, Passer domesticus domesticus (burung gereja) dan Felis maniculata domesticus (kucing jinak).</p>	<p>Langkah-langkah klasifikasi tersebut adalah sebagai berikut: mengidentifikasi objek berdasar ciri-ciri struktur tubuh makhluk hidup, misalnya, hewan atau tumbuhan yang sama jenis atau spesiesnya; setelah kelompok spesies terbentuk, dapat dibentuk kelompok-kelompok lain dari urutan tingkatan klasifikasi sebagai berikut</p>	<p>FP (2.732132178233191)</p>
<p>5.</p>	<p>Apakah yang dimaksud dengan protista?</p>	<p>Protista berasal dari bahasa Yunani protos yang berarti 'pertama'. Jadi, Protista merupakan eukariotik pertama hasil evolusi prokariotik.</p>	<p>Anggota sporozoa yang paling dikenal adalah Plasmodium</p>	<p>FP (0.9138064005898782)</p>

5.2.4 Data Hasil Pengujian *Manhattan Distance*

Setelah dilakukan pengujian menggunakan metode *Manhattan Distance* dengan kata kunci yang diperoleh pada pra-pengolahan dan perhitungan TF-IDF sebelumnya dengan nilai yang ditampilkan merupakan nilai terbaik dengan nilai terkecil yang dihitung oleh sistem dan memperoleh hasil seperti pada Tabel V-7.

Tabel V-7. Dokumen Terbaik *Manhattan Distance*

Pertanyaan	Kata Kunci Pertanyaan	Dokumen Terbaik (.txt)
Apa yang dimaksud dengan Albuminuria?	albuminuria	6
Apa yang dimaksud dengan Defekasi?	defekasi	6
Apakah yang dimaksud dengan Protoplasma ?	protoplasma	6
Apakah yang dimaksud dengan trinomial?	trinomial	6
Apakah yang dimaksud dengan protista?	protista	6

Tabel V-8. Pengujian *Manhattan Distance*

Pengujian	Confusion Matrix				Accuracy	Recall	Precision
	TP	TN	FP	FN			
Skenario 1	1	0	109	0	0.00909090909	1	0.00909090909
Skenario 2	0	0	110	0	0	0	0
Skenario 3	0	0	110	0	0	0	0
Skenario 4	0	0	110	0	0	0	0

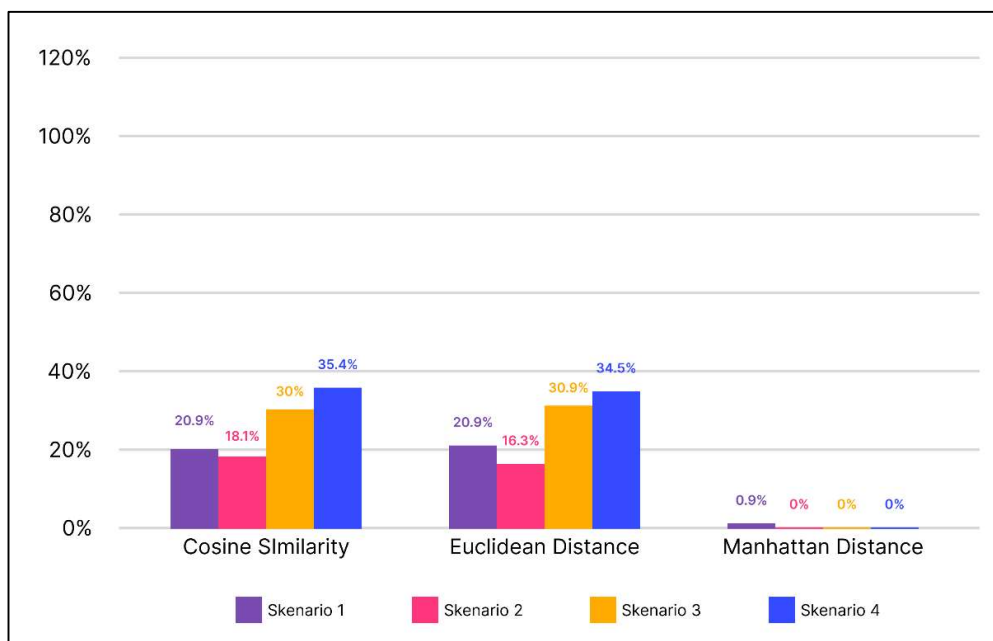
Dari hasil Tabel V-8 dapat diketahui metode *Manhattan Distance* memperoleh hasil terbaik pada skenario konfigurasi percobaan 1 dengan hasil *accuracy* dan *precision* sebesar 0.9%, serta nilai *recall* 100%. Bobot *True Positive* terbaik yang dihasilkan pada perhitungan dengan algoritma *Manhattan Distance* ini sebesar 1, sedangkan bobot *False Negative* sebesar 109. Seperti yang dijabarkan pada Tabel V-9 adalah hasil pengujian algoritma *Manhattan Distance*.

Tabel V-9. Nilai Output *Manhattan Distance*

No.	Pertanyaan	Jawaban	Keluaran Sistem	Ket.
1.	Apa yang dimaksud dengan Albuminuria?	Albuminuria, merupakan gangguan yang terjadi pada ginjal sehingga urine mengandung protein. Di dalam urine normal, sebenarnya tidak mengandung senyawa protein, asam amino, ataupun glukosa. Oleh karena itu, gangguan ini menunjukkan bila alat filtrasi pada ginjal telah rusak.	Sifat perubahan tersebut yaitu perubahan yang bersifat sementara dan perubahan yang bersifat tetap	FP (1.130333768495006)
2.	Apa yang dimaksud dengan Defekasi?	Defekasi merupakan proses pengeluaran sisa-sisa pencernaan makanan yang disebut feces melalui anus.	Sifat perubahan tersebut yaitu perubahan yang bersifat sementara dan perubahan yang bersifat tetap	FP (1.130333768495006)

3.	Apakah yang dimaksud dengan Protoplasma ?	Protoplasma adalah sejenis substansi kompleks seperti agar-agar yang tidak habis digunakan saat aktivitas kimiawi dalam menjaga kelangsungan hidup sel.	Sifat perubahan tersebut yaitu perubahan yang bersifat sementara dan perubahan yang bersifat tetap	FP (1.130333768495006)
4.	Apakah yang dimaksud dengan trinomial?	Sistem penamaan yang terdiri atas tiga suku kata disebut Trinomial, contohnya, Passer domesticus domesticus (burung gereja) dan Felis maniculata domesticus (kucing jinak).	Sifat perubahan tersebut yaitu perubahan yang bersifat sementara dan perubahan yang bersifat tetap	FP (1.130333768495006)
5.	Apakah yang dimaksud dengan protista?	Protista berasal dari bahasa Yunani protos yang berarti 'pertama'. Jadi, Protista merupakan eukariotik pertama hasil evolusi prokariotik.	Sifat perubahan tersebut yaitu perubahan yang bersifat sementara dan perubahan yang bersifat tetap	FP (1.130333768495006)

5.3 Analisis Hasil Penelitian



Gambar V- 1. Grafik Akurasi Skenario Percobaan

Pada Gambar V-1 memaparkan hasil akurasi pada setiap skenario konfigurasi percobaan yang dilakukan pada tiap algoritma pendekatan jarak. Berdasarkan pengujian yang telah dilakukan pada data uji, dapat dilihat pada Tabel V-10 bahwa algoritma yang cukup baik dalam melakukan klasifikasi pada penelitian ini adalah *Cosine Similarity* dengan skenario menggunakan *threshold* 40% jawaban keluaran sistem terhadap jawaban sebenarnya.

Tabel V-10. Akurasi Algoritma dengan Skenario Terbaik

	<i>Accuracy (%)</i>	Skenario Terbaik
<i>Cosine Similarity</i>	37.27	Skenario Percobaan 4
<i>Euclidean Distance</i>	36.36	Skenario Percobaan 4
<i>Manhattan Distance</i>	0.9	Skenario Percobaan 1

Adapun analisis yang dapat dihasilkan dari penelitian ini adalah sebagai berikut:

1. Algoritma *Euclidean Distance* dan *Cosine Similarity* mendapatkan nilai akurasi kurang baik tidak menyentuh 50% karena dipengaruhi oleh hasil perkalian dari vektor TF-IDF antara pertanyaan dan keseluruhan dokumen serta jawaban. Hal ini menyebabkan nilai terbaik pada kedua algoritma ini tidak cukup tepat untuk memperoleh bobot *True Positive* (TP) yang besar.
2. Algoritma *Manhattan Distance* memiliki akurasi kebenaran pertanyaan terhadap jawaban yang lebih rendah dikarenakan hasil perhitungan yang menghasilkan *True Positive* (TP) lebih sedikit dibandingkan algoritma lainnya. Algoritma ini menghitung selisih nilai absolut sehingga nilai terkecil yang dihasilkan tergantung kepada panjang dokumen serta jawaban yang akan dihitung.
3. Nilai akurasi pada penelitian dengan algoritma *Manhattan Distance*, *Euclidean Distance*, dan *Cosine Similarity* memiliki nilai yang cukup rendah yang disebabkan oleh 3 faktor. Pertama, pada pertanyaan dan jawaban yang dikeluarkan sistem hanya mengandalkan kata kunci yang telah melewati tahap pra-pengolahan. Kedua, pada *document retrieval* hanya mengambil dokumen dengan urutan pertama yang dianggap paling mirip dan mendekati kata kunci. Ketiga, kata kunci yang dihasilkan pada pertanyaan dan kalimat jawaban rata-rata hanya memiliki satu kata kunci dan pada kalimat kandidat jawaban kata kunci bisa muncul lebih banyak dibandingkan kata kunci pada jawaban sebenarnya yang mempengaruhi perhitungan kedekatan jarak.

4. Untuk nilai akurasi dan nilai presisi memiliki jumlah yang sama karena pada perhitungan nilai akurasi, *True Negative* (TN) tidak memiliki bobot nilai sehingga perhitungan yang hitung akan sama dengan perhitungan nilai presisi.
5. Metode pengujian yang digunakan sangat berpengaruh dalam perhitungan hasil dari nilai akurasi, nilai presisi dan nilai *recall*. Pada pengujian ini hanya akan mengeluarkan nilai *True Positive* (TP) dan *False Positive* (FP) karena jawaban yang dikeluarkan sistem selalu bernilai positif, yaitu selalu berasal dari dataset yang tersedia.

5.4 Kesimpulan

Berdasarkan pengujian model menggunakan algoritma *Manhattan Distance*, *Euclidean Distance*, dan *Cosine Similarity* pada pada Sistem Tanya Jawab untuk Pertanyaan Apa *Non-Factoid* dapat disimpulkan bahwa dari empat konfigurasi percobaan dalam pengujian, algoritma *Cosine Similarity* adalah algoritma terbaik dalam melakukan pencarian jawaban pada penelitian ini dibandingkan dengan algoritma *Manhattan Distance*.

BAB VI

KESIMPULAN DAN SARAN

6.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan, pada bab hasil dan analisis penelitian, dapat ditarik kesimpulan bahwa:

1. Sistem Tanya Jawab untuk Pertanyaan Apa Menggunakan Metode Pendekatan Jarak berhasil dikembangkan dengan membandingkan 3 algoritma yaitu, *Manhattan Distance*, *Euclidean Distance*, dan *Cosine Similarity*.
2. Dari 110 pasang pertanyaan jawaban pada dataset dan 61 dokumen menghasilkan nilai terbaik pada konfigurasi percobaan 4 dengan nilai akurasi dan nilai presisi sebesar 37.27%, serta memiliki nilai *recall* sebesar 100% pada algoritma *Cosine Similarity*. *Euclidean Distance* memiliki nilai terbaik pada konfigurasi percobaan 4 dengan nilai akurasi dan nilai presisi sebesar 36.36% dan nilai *recall* 100%. Sedangkan, pada algoritma *Manhattan Distance* nilai terbaik didapat pada konfigurasi percobaan 1 dengan nilai akurasi dan nilai presisi sebesar 0.9%, serta memiliki nilai *recall* sebesar 100%. Dari hasil tersebut dapat disimpulkan bahwa akurasi terbaik terdapat pada algoritma *Cosine Similarity* yaitu sebesar 37.27%, kemudian algoritma *Euclidean Distance* sebesar 36.36%, dan yang terkecil adalah *Manhattan Distance* yaitu sebesar 0.9%.

6.2 Saran

Saran berdasarkan penelitian yang telah dilakukan terhadap penelitian selanjutnya adalah sebagai berikut:

1. Perangkat lunak pada penelitian ini bisa dikembangkan dengan menerapkan algoritma klasifikasi lainnya yang dapat menghitung kedekatan jarak.
2. *Website* yang dikembangkan dapat digunakan oleh kelompok yang ingin mengajukan pertanyaan ‘apa’ *non-factoid* untuk memperoleh jawaban definisi tentang pembelajaran biologi.
3. Hasil penelitian mendapatkan nilai akurasi kurang baik sehingga dapat dilakukan penambahan dataset jawaban untuk pertanyaan ‘apa’ *non-factoid* agar memperoleh nilai akurasi yang lebih tinggi.
4. Pada penelitian selanjutnya dapat menggunakan metode pengujian lain untuk memperoleh nilai akurasi yang lebih besar.
5. Pada penelitian selanjutnya dapat menggunakan *Expected Answer Type* (EAT) lainnya selain *Expected Answer Type* (EAT) definisi.

DAFTAR PUSTAKA

- Almiman, A., Osman, N., & Torki, M. (2020). Deep neural network approach for arabic community question answering. *Alexandria Engineering Journal*, 59(6), 4427–4434. <https://doi.org/10.1016/j.aej.2020.07.048>
- Baskaran, U., & Ramanujam, K. (2018). Automated scraping of structured data records from health discussion forums using semantic analysis. *Informatics in Medicine Unlocked*, 10, 149–158. <https://doi.org/10.1016/j.imu.2018.01.003>
- Bouziane, A., Bouchiha, D., Doumi, N., & Malki, M. (2015). Question Answering Systems: Survey and Trends. *Procedia Computer Science*, 73, 366–375. <https://doi.org/10.1016/j.procs.2015.12.005>
- Cai, L., Zhou, S., Yan, X., & Yuan, R. (2019). A Stacked BiLSTM Neural Network Based on Coattention Mechanism for Question Answering. *Computational Intelligence and Neuroscience*, 2019. <https://doi.org/10.1155/2019/9543490>
- Casteren, W. van. (2017). *The Waterfall Model and the Agile Methodologies : A comparison by project characteristics The Waterfall Model and the Agile Methodologies : A comparison by project characteristics Academic Competences in the Bachelor 2 assignment: Write a scientific article on 2 Software Development Models*. <https://doi.org/10.13140/RG.2.2.36825.72805>
- Draxl, V. (2018). *Web Scraping Data Extraction from websites*.

- Firdaus, A., Ernawari, & Vatesia, A. (2014). APLIKASI PENDETEKSI KEMIRIPAN PADA DOKUMEN TEKS MENGGUNAKAN ALGORITMA NAZIEF & ADRIANI DAN METODE COSINE SIMILARITY. In *Jurnal Teknologi Informasi* (Vol. 10, Issue 1). <http://research.pps.dinus.ac.id>
- Fukumoto, J. (2007). *Question Answering System for Non-factoid Type Questions and Automatic Evaluation based on BE Method*. <http://clef.isti.cnr.it/>
- Fukumoto, J., Aburai, N., & Yamanishi, R. (2013). Interactive document expansion for answer extraction of question answering system. *Procedia Computer Science*, 22, 991–1000. <https://doi.org/10.1016/j.procs.2013.09.184>
- Gomaa, W. H., & Fahmy, A. A. (2013). A Survey of Text Similarity Approaches. In *International Journal of Computer Applications* (Vol. 68, Issue 13).
- Hanifah, A. F., & Kusumaningrum, R. (2021). Non-Factoid Answer Selection in Indonesian Science Question Answering System using Long Short-Term Memory (LSTM). *Procedia Computer Science*, 179, 736–746. <https://doi.org/10.1016/j.procs.2021.01.062>
- Hasnat, A., Halder, S., D, B., M, N., & D.K, B. (2013). *Comparative Study of Distance Metrics for Finding Skin Color Similarity of Two Color Facial Images*. 99–108. <https://doi.org/10.5121/csit.2013.3210>
- Jones, K. S. (1972). A statistical interpretation of term specificity and its application in retrieval. In *Journal of Documentation* (Vol. 28, Issue 1, pp. 11–21). <https://doi.org/10.1108/eb026526>

- Josi, A., & Abdillah, L. A. (2014). *PENERAPAN TEKNIK WEB SCRAPING PADA MESIN PENCARI ARTIKEL ILMIAH*.
- Kempny, C., & Brzoska, P. (2023). Anwendungskontexte von Web Scraping in der Versorgungsforschung - Nur für Web-Expert:innen? Oder eine Methode für alle Versorgungsforscher:innen!? *Zeitschrift Für Evidenz, Fortbildung Und Qualität Im Gesundheitswesen*. <https://doi.org/10.1016/j.zefq.2022.11.010>
- Kitasuka, T., Aritsugi, M., & Rahutomo, F. (2014). *Semantic Cosine Similarity*. <https://www.researchgate.net/publication/262525676>
- Kumar Sharma, S., & Kumar, S. (2016). COMPARATIVE ANALYSIS OF MANHATTAN AND EUCLIDEAN DISTANCE METRICS USING A* ALGORITHM 1 2. In *Journal of Research in Engineering and Applied Sciences JREAS* (Vol. 1). http://wiki.gamegardens.com/Path_Finding_Tut
- Li, J., Luo, Z., Huang, H., & Ding, Z. (2022). Towards Knowledge-Based Tourism Chinese Question Answering System. *Mathematics*, 10(4). <https://doi.org/10.3390/math10040664>
- Liang, M., & Niu, T. (2022). Research on Text Classification Techniques Based on Improved TF-IDF Algorithm and LSTM Inputs. *Procedia Computer Science*, 208, 460–470. <https://doi.org/10.1016/j.procs.2022.10.064>
- Patil, R., & Tiwari, S. (2020). *Question Answering System using A TF-IDF Method*. 8, 4.
- Persson, E. (2019). *Evaluating tools and techniques for web scraping*.

- Pribadi, W. W., Yunus, A., & Sartika Wiguna, A. (2022). PERBANDINGAN METODE K-MEANS EUCLIDEAN DISTANCE DAN MANHATTAN DISTANCE PADA PENENTUAN ZONASI COVID-19 DI KABUPATEN MALANG. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 6, Issue 2).
- Ryansyah A, & Andayani S. (2017). *VOLUME : 1 NO : 1 JURNAL SISTEM & TEKNOLOGI INFORMASI KOMUNIKASI* 53. <http://pdfbox.apache.org/>
- Santra, A. K., & Christy, C. J. (2012). *Genetic Algorithm and Confusion Matrix for Document Clustering*. www.IJCSI.org
- Smith, J., Petrovic, P., Rose, M., De Souza, C., Muller, L., Nowak, B., & Martinez, J. (2021). Placeholder Text: A Study. *The Journal of Citation Styles*, 3. <https://doi.org/10.10/X>
- Verberne, S., Halteren, H. van, Theijssen, D., Raaijmakers, S., & Boves, L. (2011). Learning to rank for why-question answering. *Information Retrieval*, 14(2), 107–132. <https://doi.org/10.1007/s10791-010-9136-6>
- Wahid, A. A. (2020). *Jurnal Ilmu-ilmu Informatika dan Manajemen STMIK Oktober (2020) Analisis Metode Waterfall Untuk Pengembangan Sistem Informasi*.
- Wang, J., & Dong, Y. (2020). Measurement of text similarity: A survey. In *Information (Switzerland)* (Vol. 11, Issue 9, pp. 1–17). MDPI AG. <https://doi.org/10.3390/info11090421>

- Yudiarta, N. G., Sudarma, M., & Ariastina, W. G. (2018). Penerapan Metode Clustering Text Mining Untuk Pengelompokan Berita Pada Unstructured Textual Data. *Majalah Ilmiah Teknologi Elektro*, 17(3), 339. <https://doi.org/10.24843/mite.2018.v17i03.p06>
- Yusliani, N., & Purwarianti, A. (2011). *SISTEM QUESTION ANSWERING BAHASA INDONESIA UNTUK PERTANYAAN NON-FACTOID*.
- Zeng, G. (2020). On the confusion matrix in credit scoring and its analytical properties. *Communications in Statistics - Theory and Methods*, 49(9), 2080–2093. <https://doi.org/10.1080/03610926.2019.1568485>

LAMPIRAN

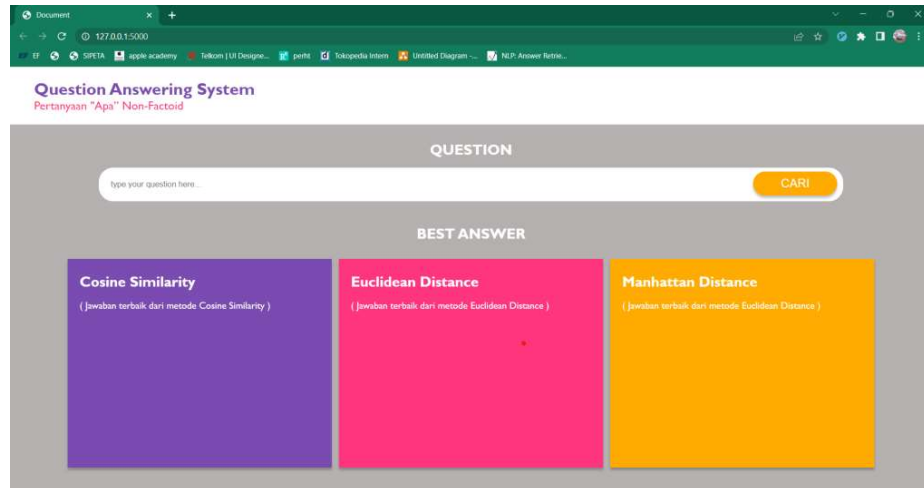
Lampiran 1. Kode Program

Pada penelitian ini, Adapun kode program serta hasil pengujian dalam mengimplementasikan *Website* Sistem Tanya Jawab terdapat pada link dibawah ini:

<https://gitlab.com/zafiragaleaskripsi/QuestionAnswerSystem.git>

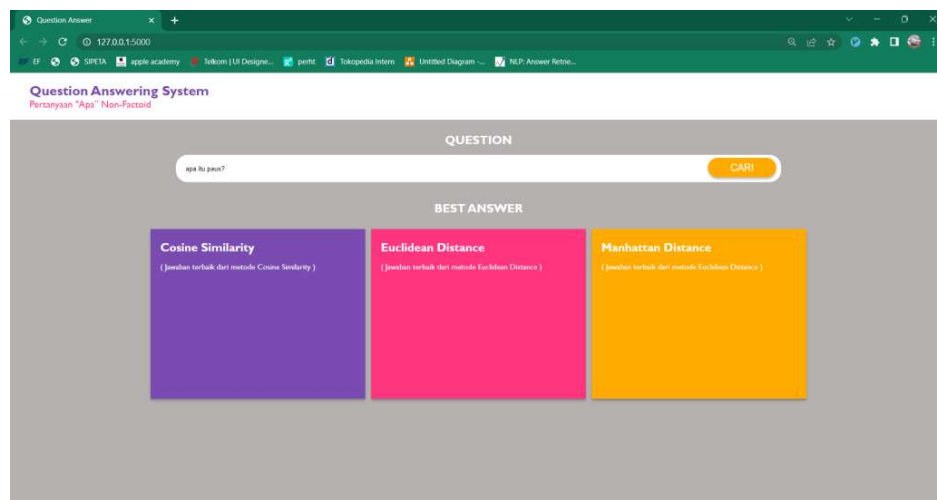
Lampiran 2. User Guide Program

Website Sistem Tanya Jawab merupakan *website* yang dapat menjawab pertanyaan apa *non-factoid* terkait pembelajaran biologi. Berikut panduan penggunaan *website* Sistem Tanya Jawab:



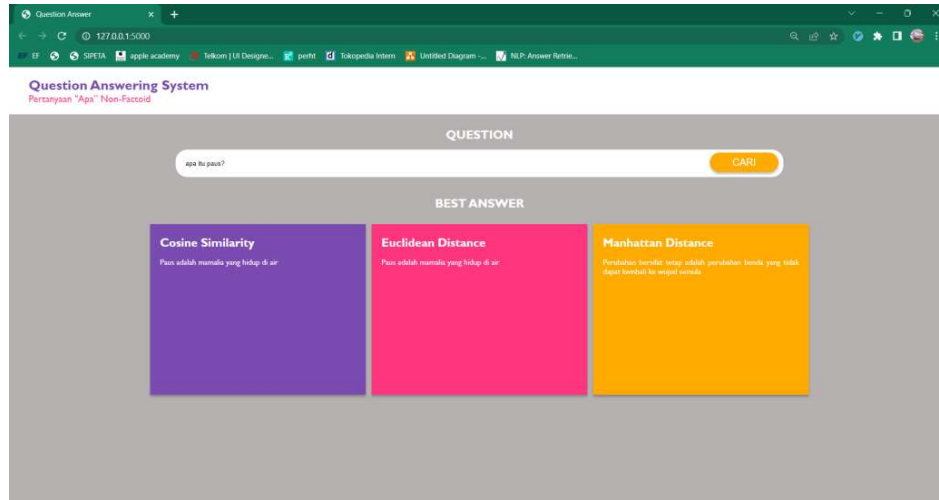
Gambar 1

- 1) Pengguna dihadapkan dengan tampilan *website* Sistem Tanya Jawab yang terdapat kolom untuk memasukkan pertanyaan.



Gambar 2

2) Pengguna dapat menekan tombol cari.



Gambar 3

3) Pengguna mendapat jawaban atas pertanyaannya dari hasil algoritma perhitungan jarak *Manhattan Distance*, *Euclidean Distance*, dan *Cosine Similarity*.

Lampiran 3. Rencana Kegiatan Penelitian

Rencana kegiatan penelitian yang dirancang pada proposal skripsi mencakupi uraian kegiatan seperti berikut:

No	Uraian Kegiatan	Tahun 2023																			
		Februari					Maret					April					Mei				
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Pengumpulan dan Analisis Data																				
a.	Analisis dan Mengumpulkan Data Pasangan Pertanyaan dan Jawaban																				
b.	Menerjemahkan Data ke Bahasa Indonesia																				
2	Membangun Sistem Tanya Jawab untuk Pertanyaan Apa Menggunakan Metode Pendekatan Jarak																				
a.	Menentukan metode pengembangan perangkat lunak																				
b.	Analisis dan perancangan arsitektur sistem QAS secara keseluruhan																				
c.	Perancangan arsitektur algoritma metode pendekatan jarak																				
d.	Menentukan kriteria pengujian																				

Lampiran 4. Sampel Data

Pada lampiran 4 ini, akan ditampilkan beberapa sampel data yang digunakan di dalam penelitian ini. Adapun sampel data tersebut dapat dilihat pada <https://gitlab.com/zafiragaleaskripsi/QuestionAnswerSystem.git> dan pada lampiran seperti berikut:

Contoh Dataset Pertanyaan dan Jawaban:

Pertanyaan	Jawaban
Apakah yang dimaksud dengan Glikoprotein ?	[Glikoprotein merupakan bagian membran sel yang tersusun atas karbohidrat dan protein.]
Apakah yang dimaksud dengan Lipoprotein ?	[Membran sel tersusun atas molekul yang disebut lipoprotein. Lipoprotein merupakan senyawa kimia yang terdiri atas lemak fosfolipid dan protein. Letak molekul lemak berada di tengah membran. Karena itu, membran ini dinamakan fosfolipid lapis ganda (bilayer fosfolipid).]
Apakah yang dimaksud dengan Lisosom ?	[Lisosom (lysis = pemisahan, pembelahan, soma = tubuh) adalah badan berbentuk bulat seperti kantong kecil dengan diameter 0,1 sampai 1 m. Di dalam lisosom terdapat 50 enzim dan kebanyakan adalah enzim hidrolitik yang bersifat asam.]
Apa yang dimaksud dengan makanan bergizi dan makanan sehat?	[Upaya memanfaatkan secara maksimal, makanan yang dikonsumsi sebaiknya mengandung berbagai zat penting (nutrisi), misalnya karbohidrat, lemak, protein, mineral, vitamin, air dan makanan berserat. Makanan yang demikian disebut makanan bergizi. Makanan juga harus mudah dicerna, higienis (bersih), gizi seimbang, dan mengandung cukup air. Makanan ini dinamakan makanan yang sehat.]
Apakah yang dimaksud dengan Osmosis ?	[Osmosis adalah perpindahan zat pelarut melalui membran selektif permeabel dari konsentrasi zat pelarut tinggi menuju konsentrasi zat pelarut rendah.]
Apa yang dimaksud dengan AIDS?	[AIDS (Acquired Immune Deficiency Syndrome). AIDS merupakan penyakit yang terjadi karena menurunnya kekebalan tubuh. AIDS disebabkan oleh infeksi virus

	<p>HIV (Human Immunodeficiency Virus). Virus ini menyerang sel darah putih yang mengakibatkan kekebalan tubuh menurun. Keberadaan HIV di dalam darah perlu waktu yang lama untuk dapat diketahui. Sebab, keberadaan HIV ini tidak memperlihatkan gejala penyakit tertentu.]</p>
<p>Apa yang dimaksud dengan penyakit emfisema paru-paru?</p>	<p>[Emfisema paru-paru merupakan suatu penyakit yang diderita seseorang akibat jumlah udara yang berlebihan di dalam paru-paru, sehingga membuat daerah pertukaran gas berkurang. Berbagai penyebab emfisema misalnya infeksi kronis oleh rokok atau bahan-bahan lain yang mengiritasi bronkus, kemudian infeksi akibat kelebihan mukus karena peradangan dan edema epitel bronkiolus. Selain itu, penyakit ini disebabkan karena adanya gangguan saluran respirasi sehingga sulit untuk berekspirasi, sehingga mengakibatkan udara terperangkap di dalam alveolus dan menyebabkan alveolus renggang.]</p>
<p>Apa yang dimaksud dengan penyakit pneumonia?</p>	<p>[Pneumonia adalah peradangan paru-paru yang mengakibatkan alveolus berisi cairan dan eritrosit secara berlebihan. Sebagai contoh pneumonia adalah pneumonia bakteri. Infeksi ini disebarkan oleh bakteri <i>Diplococcus pneumoniae</i>, dari satu alveolus ke alveolus yang lain, sehingga dapat meluas ke seluruh lobus, bahkan ke seluruh paru-paru.]</p>
<p>Apa yang dimaksud dengan Vas Deferens?</p>	<p>[Vas deferens merupakan saluran reproduksi yang berfungsi sebagai tempat Bergeraknya sperma dari epididimis menuju kantung semen (kantong mani) atau vesikula seminalis. Pada satu ujung, vas deferens menempel epididimis, sedangkan ujung lainnya berada dalam kelenjar prostat.]</p>
<p>Apa yang dimaksud dengan sembelit/konstipasi?</p>	<p>[Sembelit atau konstipasi ialah keadaan yang dialami seseorang dengan gejala feses mengeras sehingga susah dikeluarkan.]</p>

Contoh Dokumen:

Nama Dokumen	Isi Dokumen
6.txt	<p><j>Perubahan Sifat Benda.</p> <p><p>Setiap benda mempunyai sifat tertentu yang membedakannya dengan benda lain. Sifat benda meliputi bentuk, warna, kelenturan, kekerasan, dan bau. Bentuk benda bermacam-macam. Benda yang berupa bangun datar mempunyai bentuk persegi, persegi panjang, segitiga, dan lingkaran. Benda yang berupa bangun ruang mempunyai bentuk bola, kubus, balok, kerucut, dan tabung. Pelangi mempunyai warna merah, jingga, kuning, hijau, biru, nila, dan ungu. Sebagaimana pelangi, setiap benda mempunyai warna. Warna benda juga bermacam-macam. Misalnya batu berwarna hitam, mangga mentah berwarna hijau, dan jeruk matang berwarna kuning atau jingga. Kelenturan adalah sifat benda yang mudah dilengknkan. Benda yang bersifat lentur dapat dibengkokkan dan tidak mudah patah. Kekerasan adalah kemampuan suatu benda untuk menahan goresan. Suatu benda bersifat lebih keras daripada benda lain jika dapat menggores benda tersebut. Benda ada yang berbau dan ada yang tidak berbau. Bau benda meliputi harum, busuk, dan amis.</p></p> <p><p>Benda-benda dapat berubah wujud. Benda padat dapat berubah wujud menjadi benda cair ataupun gas. Demikian juga sebaliknya. Perubahan wujud ini menyebabkan perubahan sifat-sifat benda. Perubahan sifat benda meliputi bentuk, warna, kelenturan, kekerasan, dan baunya. Benda dapat berubah sifat apabila ada perlakuan atau peristiwa yang mengenainya.</p></p>

<p>Pada saat kamu memakan es krim, lama-kelamaan es krim tersebut akan mencair. Mencairnya es krim disebabkan karena suhu di luar lebih tinggi (panas) dari pada suhu es krim tersebut. Selain es krim, mentega juga mengalami hal yang sama ketika dipanaskan. Bagaimana jika air dipanaskan? Pemanasan air akan mengakibatkan air berubah wujud menjadi uap air (gas). Jadi pemanasan mengakibatkan benda mengalami perubahan wujud. Benda padat apabila dipanaskan akan berubah menjadi cair dan benda cair apabila dipanaskan akan berubah menjadi uap air.</p>

<p>Es krim atau es yang biasa kamu beli di sekolah atau warung dekat rumahmu sebenarnya berasal dari bahan-bahan yang berbentuk cairan. Apabila cairan tersebut didinginkan maka akan berubah wujud menjadi padat, yaitu es. Mentega yang dicairkan setelah dipanaskan akan kembali menjadi padat setelah didinginkan. Jadi, pendinginan menyebabkan benda mengalami perubahan wujud. Benda cair akan berubah wujudnya menjadi benda padat.</p>

<p>Pada saat kertas dibakar, kertas akan mengalami perubahan warna dan bentuk. Sebelum dibakar kertas tersebut berwarna putih, namun setelah dibakar warna kertas berubah menjadi hitam. Selain perubahan warna, kertas juga mengalami perubahan bentuk dari berupa lembaran menjadi abu. Jika kamu membakar karet maka selain bentuk dan warnanya akan berubah, kelenturan dan baunya pun menjadi berubah. Oleh karena itu, pembakaran dapat menyebabkan perubahan bentuk, warna, kelenturan, dan bau.</p>

	<p><p>Apa yang akan terjadi jika kamu menyimpan buah di udara terbuka dalam waktu beberapa hari? Tentunya buah itu akan menjadi lembek, layu, dan warnanya pun berubah. Hal ini terjadi karena buah yang dibiarkan di udara terbuka akan mengalami pembusukan. Jadi, pembusukan juga mengakibatkan benda mengalami perubahan bentuk, warna, dan bau.</p></p> <p><p>Kamu mungkin pernah melihat besi atau rantai sepedamu berkarat. Logam seperti besi, dapat mengalami perkaratan apabila terkena air atau uap air dan dibiarkan dalam waktu yang lama. Perkaratan ini menyebabkan warna besi berubah dan besi menjadi rapuh. Perkaratan dapat menyebabkan benda mengalami perubahan warna dan kekuatan.</p></p> <p><p>Pada dasarnya perubahan sifat benda dapat dibedakan menjadi dua. Sifat perubahan tersebut yaitu perubahan yang bersifat sementara dan perubahan yang bersifat tetap. Perubahan bersifat sementara adalah perubahan benda yang dapat kembali ke wujud semula dan tidak menghasilkan zat baru. Perubahan bersifat sementara disebut juga perubahan fisika. Contoh perubahan yang bersifat sementara yaitu perubahan wujud air menjadi es. Air berwujud cair, dapat berubah menjadi es yang berwujud padat. Perubahan wujud benda dari cair menjadi padat disebut membeku. Es dapat berubah wujud menjadi air kembali jika dipanaskan. Perubahan wujud ini disebut mencair. Perubahan sifat pada benda tersebut</p>
--	---

	<p>bersifat sementara, karena benda dapat kembali ke wujud semula. Perubahan bersifat tetap adalah perubahan benda yang tidak dapat kembali ke wujud semula. Perubahan ini menghasilkan zat baru. Perubahan bersifat tetap disebut juga perubahan kimia. Contoh perubahan yang bersifat tetap, yaitu perubahan wujud kertas yang dibakar menjadi abu.</p></p>
14.txt	<p><j>Dampak Peristiwa Alam. <p>Kebutuhan manusia tidak terbatas. Manusia selalu berusaha agar kebutuhan tersebut terpenuhi. Di alam telah tersedia berbagai bahan kebutuhan manusia yang disebut sumber daya alam. Sumber daya alam dapat dibedakan menjadi dua, yaitu sumber daya alam yang dapat diperbarui dan tidak dapat diperbarui. Sumber daya alam yang dapat diperbarui yaitu sumber daya alam yang selalu tersedia meskipun dimanfaatkan secara terus-menerus. Contohnya tumbuhan, hewan, air, sinar matahari, dan udara. Sumber daya alam yang tidak dapat diperbarui yaitu sumber daya alam yang akan habis jika digunakan secara terus-menerus. Sumber daya alam ini meliputi bahan tambang mineral dan nonmineral. Bahan tambang mineral contohnya aluminium, emas, perak, tembaga, nikel, dan besi. Bahan tambang nonmineral contohnya batu bara dan minyak bumi. Sumber daya alam dapat dimanfaatkan oleh manusia untuk memenuhi kebutuhannya. Namun, sangat disayangkan, terkadang manusia sampai merusak alam untuk memenuhi kebutuhannya. Perbuatan manusia inilah yang dapat mengubah permukaan bumi.</p></p>

	<p><p>Akhir-akhir ini manusia banyak melakukan pembakaran hutan untuk dijadikan lahan pertanian, permukiman penduduk, dan untuk industri. Kawasan hutan yang dijadikan lahan pertanian biasanya berubah menjadi tanah tandus dan gersang. Hal ini karena setelah panen biasanya ladang ini akan ditinggalkan. Sistem perladangan seperti ini disebut perladangan berpindah. Akhirnya hutan yang dahulu menghijau menjadi tanah tandus dan gersang. Selain untuk lahan pertanian, biasanya pembakaran hutan juga bertujuan membangun permukiman penduduk dan mendirikan pabrik.</p></p> <p><p>Selain pembakaran hutan, manusia juga melakukan penebangan hutan secara liar. Pohon-pohon ini diambil kayunya sebagai bahan bangunan. Penebangan pohon-pohon di hutan secara liar ini juga dapat mengubah permukaan bumi. Penebangan liar di Indonesia dimulai di Kalimantan pada awal tahun 1960-an. Akhirnya penebangan liar ini meluas sampai ke Sumatra dan Sulawesi. Penebangan liar ini membuat hutan di Indonesia rusak. Proses penebangan hutan secara liar disebut dengan penggundulan hutan. Pepohonan sangat penting bagi kehidupan di Bumi. Jadi, penebangan pohon harus dilakukan secara hati-hati dan disertai dengan usaha pelestariannya. Penebangan hutan harus disertai dengan penanaman kembali benih-benih pohon yang telah ditebang. Benih-benih ini akan tumbuh dan dapat menggantikan pohon-pohon yang telah ditebang. Melalui cara ini kelestarian hutan tetap terjaga. Penggundulan hutan dapat menyebabkan terjadinya perubahan</p>
--	--

<p>permukaan bumi. Hutan ini akan berubah menjadi lahan tandus dan gersang. Selain itu, penggundulan hutan juga berdampak pada kehidupan makhluk hidup. Penggundulan hutan telah membunuh ratusan ribu spesies tumbuhan dan hewan. Banyaknya pohon yang ditebangi menyebabkan hewan-hewan hutan kehilangan makanan dan tempat berlindung.</p> <p><p>Kegiatan penambangan juga dapat mengubah permukaan bumi. Sebagian besar bahan tambang berada di dalam tanah. Pengambilan bahan tambang dengan cara digali atau ditambang. Ada dua macam jenis penambangan yaitu penambangan terbuka dan penambangan bawah tanah. Penambangan terbuka adalah penambangan yang dilakukan di permukaan bumi. Beberapa bahan tambang seperti tembaga, besi, batu bara, kapur, dan aluminium sering ditemukan di permukaan bumi. Oleh karena itu, untuk mengambilnya tidak perlu menggali. Kegiatan ini mengubah bentuk permukaan bumi menjadi lubang-lubang bekas penambangan. Bahan tambang lainnya digali dari terowongan yang berada ratusan meter di bawah permukaan tanah. Cara ini disebut penambangan bawah tanah. Penambangan ini lebih sulit daripada penambangan di permukaan. Para penambang menggali sebuah lubang menuju ke dalam tanah dan mengambil bijih. Pengambilan bijih ini menggunakan bor atau bahan peledak sebelum diangkut ke permukaan. Kegiatan ini menimbulkan tanah berongga. Tanah yang berongga menyebabkan tanah kurang kuat sehingga bisa runtuh. Selain penambangan terbuka dan penambangan bawah tanah, ada juga cara lainnya yaitu</p>	
--	--

	<p>pengerukan. Pengerukan merupakan cara lain yang digunakan untuk mengumpulkan logam-logam yang terendap di dalam batuan di dasar sungai atau sumber air lainnya. </p></p>
--	---