BAB II
TINJAUAN PUSTAKA

2.1 Penelitian Yang Relevan

Untuk mengetahui perbandingan penelitian yang dilakukan dengan penelitian yang sudah ada, maka perlu dilakukan studi sejenis. Berikut merupakan beberapa penelitian terdahulu.

Penelitian yang dilakukan oleh (Fauzi et al., 2025) pada tahun 2025 dengan judul "***Aplikasi Chatbot Rekomendasi Laptop Menggunakan Natural Language Processing***" membahas pengembangan chatbot rekomendasi laptop dengan pendekatan Natural Language Processing menggunakan algoritma *Term Frequency–Inverse Document Frequency* (TF-IDF) dan *cosine similarity*. Sistem dirancang untuk memproses masukan teks dari pengguna dan menghasilkan rekomendasi laptop yang sesuai dengan kebutuhan yang diinginkan. Aplikasi dikembangkan menggunakan bahasa pemrograman *Python* dengan *framework Flask* serta basis data *MySQL*. Hasil pengujian menunjukkan bahwa sistem mampu memberikan rekomendasi dengan tingkat akurasi yang cukup baik. Namun demikian, pendekatan berbasis representasi vektor dan pembelajaran semantik ini memerlukan proses komputasi yang relatif kompleks dan kurang optimal untuk diterapkan pada dokumen teknis yang jumlahnya terbatas tetapi memiliki struktur dan terminologi yang sangat spesifik, seperti dokumen kebijakan geospasial.

Penelitian yang dilakukan oleh (Mulyatun et al., 2021) pada tahun 2021 dengan judul "***Pendekatan Natural Language Processing pada Aplikasi Chatbot sebagai Alat Bantu Customer Service***" mengkaji penerapan Natural Language Processing dengan metode *FSM* dalam pengembangan *chatbot* layanan informasi akademik. *Chatbot* dirancang untuk menjawab pertanyaan mahasiswa terkait jadwal perkuliahan, data dosen, dan administrasi akademik secara otomatis tanpa batasan waktu layanan. Metode *FSM* digunakan untuk mengidentifikasi tingkat kemiripan teks sehingga sistem tetap mampu memberikan respons yang relevan meskipun terdapat kesalahan pengetikan atau variasi kata pada input pengguna. Hasil penelitian menunjukkan bahwa sistem dapat meningkatkan efisiensi layanan informasi akademik dan membantu pengguna memperoleh informasi secara cepat dan akurat.

Penelitian yang dilakukan oleh (Dharma Putra et al., 2025) pada tahun 2025 dengan judul **"*Analisa Pencarian dan Pencocokan String dalam Aplikasi Berbasis Python Menggunakan Library fuzzywuzzy terhadap Dataset CNN Daily Mail*"** membahas implementasi *FSM* sebagai dasar pencarian cerdas dalam aplikasi berbasis teks. Penelitian ini menguji kemampuan sistem dalam mencocokkan string pada kondisi input yang tidak sempurna, seperti kesalahan pengetikan, singkatan, dan variasi ejaan. Evaluasi dilakukan menggunakan dataset *CNN Daily Mail* dengan berbagai skenario pencocokan entitas. Hasil penelitian menunjukkan bahwa kombinasi fungsi *fuzz.partial_ratio* dan *process.extractOne* mampu menghasilkan tingkat akurasi hingga 92% pada pencocokan entitas spesifik serta tetap stabil meskipun terdapat noise pada input. Temuan ini menjadi dasar teknis dalam penerapan *fuzzywuzzy* pada pencarian entitas dokumen terstruktur.

Penelitian yang dilakukan oleh (Widianto, 2022) pada tahun 2022 dengan judul **"*Rancang Bangun Aplikasi Chatbot untuk Pendukung Perdagangan dengan Metode Fuzzy String Matching Menggunakan RUP*"** membahas pengembangan chatbot pendukung perdagangan pada UMKM dengan pendekatan *Rational Unified Process* (RUP). Metode *FSM* digunakan untuk mengenali masukan pengguna secara fleksibel meskipun terjadi kesalahan pengetikan. Proses pengembangan dilakukan secara iteratif melalui tahapan inisialisasi, elaborasi, konstruksi, dan transisi. Hasil penelitian menunjukkan bahwa sistem mampu meningkatkan efisiensi pelayanan, mengurangi beban kerja manual, serta memberikan respons yang relevan dan cepat.

Penelitian yang dilakukan oleh (Ahmad Zainul Fanani et al., 2022) pada tahun 2022 dengan judul **"*Sistem Customer Service Cerdas Menggunakan Metode Fuzzy String Matching pada E-Commerce*"** membahas pengembangan chatbot layanan pelanggan pada platform e-commerce. Penelitian ini berangkat dari permasalahan kesalahan pengetikan input pengguna yang menyebabkan kegagalan pencarian pada metode exact matching. Dengan menerapkan *FSM*, sistem mampu mengukur tingkat kemiripan antara input pengguna dan basis data tanya jawab. Hasil penelitian menunjukkan bahwa sistem dapat menghasilkan respons yang valid dan relevan meskipun input tidak sempurna, serta mampu mengurangi jumlah pertanyaan yang tidak terjawab.

Penelitian yang dilakukan oleh (Rohman et al., 2023) pada tahun 2023 dengan judul **"*Chatbot untuk Cek Persediaan Stok Barang Menggunakan Metode Fuzzy String Matching Berbasis Mobile*"** merancang chatbot untuk membantu pelanggan dalam memeriksa ketersediaan dan harga barang. Metode *FSM* digunakan untuk mencocokkan kata kunci input pengguna dengan data produk dalam basis data. Hasil pengujian menunjukkan bahwa sistem mampu memberikan respons dengan tingkat akurasi yang tinggi dan waktu tanggap yang cepat, sehingga meningkatkan efisiensi pelayanan dan kemudahan akses informasi produk.

Penelitian yang dilakukan oleh (Devalio et al., 2025) pada tahun 2025 dengan judul **"*Perancangan Website Prediksi Jumlah Pengeluaran Mahasiswa Berbasis Streamlit*"** membahas pengembangan sistem prediksi pengeluaran mahasiswa menggunakan *framework Streamlit*. Sistem dikembangkan melalui tahapan pengolahan data, pelatihan model, dan pengujian. Hasil pengujian menggunakan metode *blackbox testing* menunjukkan bahwa sistem mampu menghasilkan prediksi yang stabil dan dapat digunakan sebagai alat bantu pengelolaan keuangan mahasiswa.

Penelitian yang dilakukan oleh (Henri Saputro et al., 2021) pada tahun 2021 dengan judul **"*Rancang Bangun Aplikasi Template Surat Online dengan Metode Fuzzy String Matchimng*"** mengembangkan sistem pembuatan dan pencarian surat resmi berbasis web. Metode *FSM* diterapkan untuk mempercepat pencarian surat berdasarkan kata kunci meskipun terdapat perbedaan ejaan. Hasil penelitian menunjukkan bahwa sistem mampu meningkatkan efisiensi administrasi serta konsistensi pengelolaan dokumen resmi.

Penelitian yang dilakukan oleh (Karmila et al., 2022) pada tahun 2022 dengan judul ***"Metode Latent Dirichlet Allocation untuk Menentukan Topik Teks Suatu Berita"*** membahas penerapan *Latent Dirichlet Allocation (LDA)* untuk topic modeling dalam dokumen berita dari portal berita online. Sistem ini dirancang untuk mengidentifikasi dan mengelompokkan topik yang sedang tren berdasarkan dokumen yang dianalisis. Hasil penelitian menunjukkan bahwa LDA berhasil menghasilkan tiga topik utama dengan akurasi sebesar 67%. Namun demikian, pendekatan berbasis LDA ini lebih mengarah pada pemodelan topik secara statistik yang memerlukan waktu lebih lama dalam menentukan topik utama dibandingkan dengan metode lain, seperti *FSM (FSM)*, yang

lebih cocok diterapkan pada dokumen geospasial yang memiliki struktur dan terminologi teknis yang sangat spesifik.

Penelitian yang dilakukan oleh (Asri et al., 2025) pada tahun 2025 dengan judul ***"Peningkatan Performa Ulasan Berbahasa Indonesia dengan Spelling Corrector Peter Norvig dan Pelabelan SentiStrength_ID"*** membahas penerapan spelling corrector berbasis Peter Norvig dalam meningkatkan kualitas analisis sentimen pada ulasan aplikasi PLN Mobile. Penelitian ini mengimplementasikan teknik spelling correction pada tahap preprocessing data untuk memperbaiki kesalahan ketik dalam teks ulasan pengguna. Hasil penelitian menunjukkan bahwa penggunaan spelling corrector ini dapat meningkatkan akurasi analisis sentimen secara signifikan, dengan akurasi mencapai 82%. Namun demikian, meskipun efektif untuk perbaikan ejaan, teknik ini lebih cocok untuk teks yang bersifat lebih umum, dan penerapannya pada dokumen teknis dengan istilah spesifik, seperti dokumen kebijakan geospasial, memerlukan penyesuaian agar tetap optimal dalam mendeteksi kesalahan teknis dalam penulisan istilah.

Penelitian yang dilakukan oleh (Setyo Purwanto et al., 2022) dengan judul ***"Information Retrieval In Text-Based Document Using Boyer Moore Algorithm"*** membahas penggunaan Boyer-Moore Algorithm dalam pencarian dokumen berbasis teks. Algoritma ini bertujuan untuk mempercepat proses pencocokan string dengan menggunakan dua teknik utama: looking-glass technique dan character-jump technique, yang memungkinkan pencocokan dilakukan dengan lebih efisien. Penelitian ini menunjukkan bahwa Boyer-Moore sangat efektif dalam meningkatkan kecepatan pencarian dokumen. Namun demikian, pendekatan ini lebih tepat digunakan dalam pencocokan exact match dan tidak memiliki toleransi terhadap kesalahan penulisan, yang membuatnya kurang optimal untuk diterapkan pada dokumen teknis seperti dokumen kebijakan geospasial yang sering kali memiliki variasi penulisan dan kesalahan ketik.

Penelitian yang dilakukan oleh (Kusuma et al., 2024) pada tahun 2024 dengan judul ***"An Analysis of Electrical Energy Usage in Social Consumers Using Fuzzy C Means"*** membahas penerapan Fuzzy C Means (FCM) dalam analisis pola konsumsi energi listrik di konsumen sosial, seperti tempat ibadah, lembaga pendidikan, dan layanan kesehatan. FCM digunakan untuk mengelompokkan data berdasarkan pola konsumsi yang serupa,

meskipun data tersebut memiliki fluktuasi tinggi. Hasil penelitian menunjukkan bahwa FCM dapat mengidentifikasi tiga kelompok utama berdasarkan pola konsumsi listrik, yaitu konsumsi rendah, menengah, dan tinggi. Meskipun konteksnya berbeda, yaitu untuk pola konsumsi energi, penggunaan FCM untuk pengelompokan data berbasis pola ini dapat diterapkan dalam sistem pencarian dokumen geospasial yang mengelompokkan dokumen berdasarkan kategori atau entitas tertentu, seperti yang dilakukan dalam penelitian ini.

Penelitian yang dilakukan oleh (Dody et al., 2023) pada tahun 2023 dengan judul ***"Web-Based Knowledge Management System Application to Improve Employee Activities"*** membahas perancangan dan pengembangan sistem manajemen pengetahuan berbasis web untuk mendukung aktivitas kerja pegawai. Penelitian ini menekankan pentingnya sistem informasi yang mampu mengelola, menyimpan, dan menelusuri informasi secara terstruktur agar mudah diakses oleh pengguna. Evaluasi sistem dilakukan melalui pengujian fungsional untuk memastikan setiap fitur berjalan sesuai kebutuhan pengguna. Hasil penelitian menunjukkan bahwa sistem manajemen pengetahuan berbasis web mampu meningkatkan efisiensi kerja serta mempermudah akses informasi internal. Penelitian ini relevan dengan penelitian yang dilakukan karena sama-sama menitikberatkan pada pengembangan sistem informasi berbasis web dan pengujian fungsional sistem, khususnya dalam konteks pencarian dan penyajian informasi.

Penelitian yang dilakukan oleh (Mardiana et al., 2022) pada tahun 2022 dengan judul "***Sistem Navigasi Augmented Reality dengan Pencarian Jalur Terbaik Menuju Lokasi Pustaka (Studi Kasus pada UPT Perpustakaan UNILA)***" membahas pengembangan sistem pencarian lokasi dalam gedung menggunakan teknologi Augmented Reality (AR). Sistem ini dirancang untuk memberikan informasi berupa jarak, waktu, dan rute yang dilalui untuk sampai ke lokasi tujuan, khususnya di Perpustakaan Universitas Lampung. Penelitian ini menguji tiga algoritma pencarian jalur, yaitu *A Star (A)**, Dijkstra, dan Best First Search (BFS), untuk menentukan jalur tercepat. Hasil penelitian menunjukkan bahwa Algoritma Dijkstra memberikan hasil terbaik dengan tingkat akurasi 90%, dibandingkan dengan A Star (A)* yang mencapai 70%, dan BFS yang tidak menghasilkan

jalur yang sesuai. Sistem ini menunjukkan keberhasilan dalam memberikan navigasi berbasis AR dan dapat mempercepat pencarian lokasi di gedung besar seperti perpustakaan.

Penelitian yang dilakukan oleh (Karyadi et al., 2024) pada tahun 2024 dengan judul "***Sistem Navigasi dan Rekomendasi Buku Perpustakaan Berbasis Augmented Reality***" membahas pengembangan sistem navigasi dan rekomendasi buku menggunakan teknologi Augmented Reality (AR) di Perpustakaan Universitas Pradita. Sistem ini dirancang untuk membantu pengguna mencari buku serta memberikan rekomendasi buku sejenis, menggunakan metode markerless AR. Algoritma A* digunakan untuk navigasi pencarian jalur terpendek, sementara TF-IDF dan Cosine Similarity digunakan dalam sistem rekomendasi buku. Hasil pengujian menunjukkan bahwa aplikasi ini berhasil memudahkan pencarian buku dan rekomendasi buku sejenis tanpa memerlukan bantuan petugas, serta memperoleh tingkat penerimaan pengguna yang baik dengan skor 78,42 pada System Usability Scale. Temuan ini relevan untuk penelitian sistem informasi pencarian dalam dokumen geospasial karena pendekatan AR dan sistem rekomendasi berbasis teks dapat diadaptasi untuk meningkatkan pencarian informasi dalam dokumen teknis geospasial yang kompleks.

Penelitian yang dilakukan oleh (Ardiansyah et al., 2019) dengan judul "***Sistem Informasi Penjualan Perlengkapan Tidur (SIPPAT) Berbasis Web pada Fortun Barokah Karawang***" membahas pengembangan sistem informasi berbasis web untuk penjualan perlengkapan tidur. Sistem ini dirancang untuk mempermudah pencarian produk, transaksi pembelian, serta memberikan kemudahan akses informasi kepada pelanggan. Penelitian ini menggunakan Entity Relationship Diagram (ERD) dan Logical Record Structure (LRS) untuk merancang struktur sistem dan basis data. Hasil penelitian menunjukkan bahwa sistem berbasis web ini meningkatkan efisiensi dalam pengelolaan data produk dan transaksi, serta membantu mempermudah promosi produk secara online. Penelitian ini dapat dijadikan referensi untuk pengembangan sistem informasi pencarian pada dokumen yang memiliki struktur dan informasi yang kompleks, dengan mengadaptasi sistem navigasi berbasis web untuk pencarian entitas dan pengelolaan data secara lebih efisien.

Penelitian yang dilakukan oleh (Syaumi et al., 2025) pada tahun 2025 dengan judul "***Sistem Informasi Navigasi Wisata Kota Jakarta untuk Menentukan Rute Tercepat Menggunakan Algoritma Dijkstra Berbasis Web***" membahas pengembangan sistem navigasi wisata berbasis web untuk menentukan rute tercepat di Jakarta menggunakan algoritma Dijkstra. Sistem ini dirancang untuk membantu wisatawan mencari jalur tercepat antara lokasi asal dan tujuan dengan memanfaatkan data koordinat geografis dan jarak antar lokasi. Dengan penerapan algoritma Dijkstra, sistem menghitung rute optimal dan estimasi waktu perjalanan, serta menampilkan rute tercepat secara real-time. Penelitian ini relevan dengan pengembangan sistem informasi pencarian berbasis pencarian entitas dalam dokumen geospasial, karena menggunakan algoritma pencarian rute yang efisien dan dapat diadaptasi untuk pencarian informasi kunci dalam dokumen geospasial yang kompleks, seperti LCODE, IGT, dan atribut terkait.

Penelitian yang dilakukan oleh (Suhendra Anjar Dinata et al., 2026) pada tahun 2025 dengan judul "***Navigasi Digital Aman: Edukasi Teknik Penelusuran Informasi (Advanced Search) dan Mitigasi Risiko Unduhan di Web Terbuka***" membahas pengembangan program pelatihan untuk meningkatkan literasi informasi dan keamanan digital di kalangan siswa SMK IT Yasiba. Penelitian ini menggunakan pendekatan Participatory Action Research (PAR) untuk mengajarkan teknik pencarian informasi tingkat lanjut (Advanced Search) dengan operator logika mesin pencari, serta mitigasi risiko unduhan berbahaya dari web terbuka. Hasil penelitian menunjukkan bahwa pelatihan ini meningkatkan efisiensi pencarian informasi sebesar 85% dan kesadaran terhadap ancaman siber berbasis unduhan sebesar 78%, serta menciptakan budaya navigasi digital yang lebih aman dan cerdas di kalangan siswa. Penelitian ini relevan untuk sistem informasi pencarian, karena menggabungkan teknik pencarian tingkat lanjut yang bisa diadaptasi untuk memudahkan pencarian data/entitas dalam dokumen geospasial yang kompleks.

2.1.1 Matrix Penelitian

Tabel 2. 1 Matrix Penelitian

No	Nama Peneliti	Judul Artikel	Metode	Kriteria	Hasil Penelitian	Keterbatasan	Kesimpulan
1.	(Fauzi et al., 2025)	Aplikasi Chatbot Rekomendasi Laptop Menggunakan Natural Language Processing	TF-IDF, Cosine Similarity	Spesifikasi laptop	Sistem mampu memberikan rekomendasi dengan akurasi yang baik	Komputasi relatif kompleks, kurang sesuai untuk dokumen teknis terbatas	NLP berbasis vektor efektif untuk rekomendasi, tetapi tidak optimal untuk dokumen teknis spesifik
2.	(Mulyatun et al., 2021)	Pendekatan Natural Language Processing pada Aplikasi Chatbot Sebagai Alat Bantu Customer Service	FSM	Informasi akademik	FSM efektif meningkatkan efisiensi pencarian dan respons chatbot	Terbatas pada domain akademik	FSM cocok untuk pencarian teks toleran kesalahan input
3.	(Dharma Putra et al., 2025)	Analisa Pencarian dan Pencocokan String dalam Aplikasi Berbasis Python dengan Library FuzzyWuzzy terhadap	dan Pencocokan String Menggunakan FuzzyWuzzy FuzzyWuzzy (Levenshtein)	Dataset CNN Daily Mail	Akurasi, presisi, recall, dan F1-score mencapai 1.0	Dataset tidak berupa dokumen kebijakan	FuzzyWuzzy sangat efektif untuk pencocokan string tidak eksak

No	Nama Peneliti	Judul Artikel	Metode	Kriteria	Hasil Penelitian	Keterbatasan	Kesimpulan
		Dataset CNN Daily Mail					
4.	(Widianto, 2022)	Rancang Bangun Aplikasi Chatbot untuk Pendukung Perdagangan dengan Menggunakan Metode FSM–RUP	FSM, RUP	Informasi produk UMKM	Chatbot meningkatkan efisiensi layanan dan respons sistem	Skala UMKM, data terbatas	FSM efektif diterapkan pada chatbot berbasis dokumen terstruktur
5.	(Ahmad Zainul Fanani et al., 2022)	Sistem Customer Service Cerdas Menggunakan Metode FSM pada E-Commerce	FSM	Database tanya–jawab yang diinput manual	Sistem mampu menangani kesalahan ketik dan meningkatkan efisiensi layanan	Bergantung pada kualitas basis data	FSM meningkatkan akurasi pencarian pada layanan berbasis teks
6.	(Rohman et al., 2023)	Chatbot untuk Cek Persediaan Stok Barang Menggunakan Metode FSM Berbasis Mobile	FSM	Data produk dan harga	Pola pencarian string menggunakan FSM	Fokus pada pencarian sederhana berbasis produk	FSM efisien untuk pencarian teks cepat pada sistem mobile
7.	(Devalio et al., 2025)	Perancangan Website Prediksi Jumlah Pengeluaran Mahasiswa Berbasis	Streamlit, CatBoost	Data pengeluaran mahasiswa	Sistem prediksi berjalan stabil dan fungsional	Fokus pada prediksi, bukan pencarian teks	Streamlit efektif sebagai antarmuka web ringan dan interaktif untuk

No	Nama Peneliti	Judul Artikel	Metode	Kriteria	Hasil Penelitian	Keterbatasan	Kesimpulan
		Framework Streamlit					menampilkan hasil pencarian berbasis FSM.
8.	(Henri Saputro et al., 2021a)	Rancang Bangun Aplikasi Template Surat Online dengan Metode FSM	FSM	Dokumen surat resmi	Pencarian surat lebih cepat dan akurat	Terbatas pada dokumen akademik	FSM efektif untuk pencarian dokumen administratif
9.	(Karmila et al., 2022)	Metode Latent Dirichlet Allocation untuk Menentukan Topik Teks Suatu Berita	Latent Dirichlet Allocation (LDA)	Akurasinya, pemodelan topik	LDA berhasil menghasilkan tiga topik utama dengan akurasi 67%.	Fokus pada pemodelan topik, bukan pencarian teks.	Penelitian ini relevan karena sama-sama menggunakan pendekatan pemodelan topik untuk mengelompokkan informasi. Meskipun LDA lebih cocok untuk topik modeling berbasis dokumen, FSM dapat memberikan hasil lebih cepat dengan toleransi

No	Nama Peneliti	Judul Artikel	Metode	Kriteria	Hasil Penelitian	Keterbatasan	Kesimpulan
							kesalahan penulisan, yang berguna dalam sistem pencarian dokumen geospasial.
10.	(Asri et al., 2025)	Peningkatan Performa Ulasan Berbahasa Indonesia dengan Spelling Corrector Peter Norvig dan Pelabelan SentiStrength_ID	Spelling Corrector Peter Norvig dan SentiStrength_ID	Akurasi koreksi ejaan, akurasi sentimen	Penggunaan spelling corrector berbasis Peter Norvig berhasil meningkatkan akurasi analisis sentimen hingga 82%.	Fokus pada perbaikan ejaan, bukan pencarian teks.	Penelitian ini relevan dengan penelitian penulis karena spelling correction sangat penting dalam mengatasi kesalahan penulisan yang sering terjadi dalam dokumen geospasial. Penerapan teknik spelling correction ini pada tahap preprocessing dapat meningkatkan kualitas pencocokan

No	Nama Peneliti	Judul Artikel	Metode	Kriteria	Hasil Penelitian	Keterbatasan	Kesimpulan
							string dalam sistem FSM yang penulis kembangkan .
11.	(Setyo Purwanto et al., 2022)	Information Retrieval In Text-Based Document Using Boyer Moore Algorithm	Boyer–Moore	Kecepatan pencocokan string, efisiensi	Boyer-Moore meningkatkan kecepatan pencarian dokumen secara signifikan, namun hanya efektif untuk exact match.	Terbatas pada exact match, tidak toleran terhadap kesalahan penulisan .	Penelitian ini relevan karena Boyer-Moore sangat efektif dalam pencocokan kata yang identik, namun kurang toleran terhadap kesalahan penulisan. Dalam konteks dokumen geospasial, FSM lebih cocok karena dapat menangani kesalahan penulisan, variasi istilah, dan singkatan yang sering

No	Nama Peneliti	Judul Artikel	Metode	Kriteria	Hasil Penelitian	Keterbatasan	Kesimpulan
							ada pada dokumen teknis.
12.	(Kusuma et al., 2024)	An Analysis of Electrical Energy Usage in Social Consumers Using Fuzzy C Means	Fuzzy C Means (FCM)	Pola pengelompokan data berbasis pola	FCM dapat mengelompokkan data berdasarkan pola konsumsi energi listrik, menghasilkan tiga cluster utama: konsumsi rendah, menengah, dan tinggi.	Terbatas pada data energi, bukan pencarian teks.	FCM relevan karena menggunakan fuzzy logic untuk mengelompokkan data berdasarkan pola. Konsep ini bisa diterapkan dalam pengelompokan dokumen geospasial yang mengelompokkan dokumen berdasarkan kategori atau entitas tertentu, sesuai dengan FSM yang digunakan dalam pencarian dokumen.

No	Nama Peneliti	Judul Artikel	Metode	Kriteria	Hasil Penelitian	Keterbatasan	Kesimpulan
13.	(Dody et al., 2023)	Web-Based Knowledge Management System Application to Improve Employee Activities	Sistem Informasi Berbasis Web	Fungsionalitas sistem, kemudahan akses informasi	Sistem berbasis web meningkatkan efisiensi pengelolaan dan pencarian informasi internal organisasi	Fokus pada manajemen pengetahuan, bukan pencarian teks.	Penelitian ini relevan karena mengembangkan sistem berbasis web untuk manajemen pengetahuan. FSM dapat diintegrasikan dalam sistem tersebut untuk meningkatkan pencarian dokumen geospasial, dengan memastikan sistem lebih user-friendly dan lebih efisien dalam mengelola informasi.
14	(Mardiana et al., 2022)	Sistem Navigasi Augmented Reality dengan Pencarian Jalur Terbaik Menuju	A Star (A*), Dijkstra, BFS	Pencarian jalur terpendek	Algoritma Dijkstra memberikan hasil terbaik dengan akurasi 90%, dibandingkan dengan A	Hanya diuji pada perpustakaan, kurang efektif di lingkungan yang	Sistem informasi pencarian cocok untuk mencari data berdasarkan jalur terpendek,

No	Nama Peneliti	Judul Artikel	Metode	Kriteria	Hasil Penelitian	Keterbatasan	Kesimpulan
		Lokasi Pustaka (Studi Kasus pada UPT Perpustakaan UNILA)			Star (A) 70%, dan BFS yang tidak menghasilkan jalur sesuai	lebih kompleks	yang dapat diadaptasi untuk pencarian entitas dalam dokumen geospasial tanpa perlu membuka dokumen fisik.
15	(Karyadi et al., 2024)	Sistem Navigasi dan Rekomendasi Buku Perpustakaan Berbasis Augmented Reality	Markerless AR, A*	Pencarian jalur terpendek	Algoritma A* digunakan untuk navigasi jalur terpendek, TF-IDF dan Cosine Similarity untuk rekomendasi buku. Aplikasi berhasil meningkatkan efisiensi pencarian buku dan rekomendasi, dengan skor 78,42 pada System	Terbatas pada penggunaan di perpustakaan, kurang diuji pada dokumen lain yang lebih kompleks	Sistem informasi pencarian berbasis AR berhasil meningkatkan efisiensi pencarian buku dan rekomendasi dan rekomendasi berbasis teks dapat diterapkan pada pencarian informasi dalam dokumen geospasial, mempermudah pengguna menemukan

No	Nama Peneliti	Judul Artikel	Metode	Kriteria	Hasil Penelitian	Keterbatasan	Kesimpulan
					Usability Scale		data sekaligus meningkatkan efisiensi pencarian.
16	(Syaumi et al., 2025)	Sistem Informasi Navigasi Wisata Kota Jakarta untuk Menentukan Rute Tercepat Menggunakan Algoritma Dijkstra Berbasis Web	Dijkstra	Pencarian rute tercepat	Sistem dapat menghitung rute optimal dan estimasi waktu perjalanan dengan menggunakan algoritma Dijkstra, serta menampilkan rute tercepat secara real-time. Hasil penelitian menunjukkan bahwa Dijkstra memberikan rute tercepat dengan akurasi tinggi.	Terbatas pada pengujian di Jakarta dan tidak diuji pada konteks pencarian informasi dalam dokumen geospasial	Sistem informasi pencarian berbasis algoritma Dijkstra menampilkan rute tercepat secara real-time dengan akurasi tinggi. Sistem informasi pencarian dapat diadaptasi untuk pencarian informasi kunci dalam dokumen geospasial yang kompleks, seperti LCODE, IGT, dan atribut

No	Nama Peneliti	Judul Artikel	Metode	Kriteria	Hasil Penelitian	Keterbatasan	Kesimpulan
							terkait, dengan meningkatkan efisiensi pencarian data dan keakuratan rujukan dalam dokumen.
17	(Ardiansyah et al., 2019)	Sistem Informasi Penjualan Perlengkapan Tidur (SIPPAT) Berbasis Web pada Fortun Barokah Karawang	ERD, LRS, Web-based	Pencarian produk, transaksi pembelian	Sistem berbasis web berhasil meningkatkan efisiensi dalam pengelolaan data produk dan transaksi serta mempermudah promosi produk online	Terbatas pada aplikasi penjualan dan tidak diuji pada dokumen atau data yang lebih kompleks	Sistem berbasis web dapat diterapkan untuk navigasi informasi dalam dokumen yang memiliki struktur kompleks, membantu pencarian entitas dan pengelolaan data secara lebih efisien.
18	(Suhendra Anjar Dinata et al., 2026)	Navigasi Digital Aman: Edukasi Teknik Penelusuran Informasi	Participatory Action Research (PAR)	Pencarian informasi tingkat lanjut	Penelitian ini menunjukkan bahwa pelatihan teknik pencarian	Terbatas pada edukasi di kalangan siswa	Teknik pencarian tingkat lanjut Advanced Search dapat diadaptasi

No	Nama Peneliti	Judul Artikel	Metode	Kriteria	Hasil Penelitian	Keterbatasan	Kesimpulan
		(Advanced Search) dan Mitigasi Risiko Unduhan di Web Terbuka			informasi dengan operator logika mesin pencari meningkatkan efisiensi pencarian sebesar 85% dan kesadaran terhadap ancaman siber berbasis unduhan sebesar 78%	SMK IT dan hanya diuji pada web terbuka	untuk mempermudah pencarian entitas dalam dokumen geospasial yang kompleks, serta meningkatkan efisiensi dan keamanan pencarian data dalam dokumen berbasis teks.

Berdasarkan hasil kajian penelitian terdahulu, penulis mengidentifikasi beberapa kesenjangan penelitian yang menjadi dasar dilakukannya penelitian ini.

1. Penelitian sebelumnya lebih banyak menggunakan dataset dokumen administratif atau percakapan chatbot yang memiliki struktur sederhana. Penelitian ini mengisi kesenjangan dengan fokus pada dokumen geospasial yang lebih teknis dan kompleks, seperti *LCODE* dan IGT yang memerlukan pencocokan meskipun ada kesalahan penulisan atau variasi istilah.
2. Sebagian besar penelitian menggunakan metode FSM untuk pencocokan string pada data teks sederhana. Penelitian ini mengadaptasi metode tersebut pada dokumen geospasial yang lebih kompleks dengan menambahkan teknik normalisasi dan threshold untuk meningkatkan kualitas pencarian meskipun ada variasi ejaan atau kesalahan ketik.
3. Banyak penelitian sebelumnya mengembangkan sistem pencocokan berbasis web dengan kebutuhan perangkat keras yang tinggi. Penelitian ini mengembangkan

sistem berbasis Streamlit yang lebih praktis dan tidak bergantung pada perangkat keras khusus, membuatnya lebih mudah diakses dan diimplementasikan di lingkungan Kementerian Kehutanan.

4. Penelitian terdahulu lebih banyak mengandalkan akurasi sebagai satu-satunya metrik untuk mengukur pencocokan string. Penelitian ini menambahkan skor kemiripan dan threshold untuk memastikan hanya hasil yang memenuhi kriteria relevansi yang ditampilkan, sehingga kualitas pencarian lebih terjamin.
5. Penelitian ini berfokus pada penerapan sistem pencarian berbasis navigasi informasi, yang memudahkan pencarian entitas kunci dalam dokumen geospasial seperti LCODE, IGT, dan atribut terkait. Sistem ini menggunakan FSM untuk meningkatkan akurasi dan kecepatan pencarian, serta memastikan hasil yang relevan meskipun terdapat kesalahan ketik atau variasi istilah dalam dokumen yang teknis.

2.2 Landasan Teori

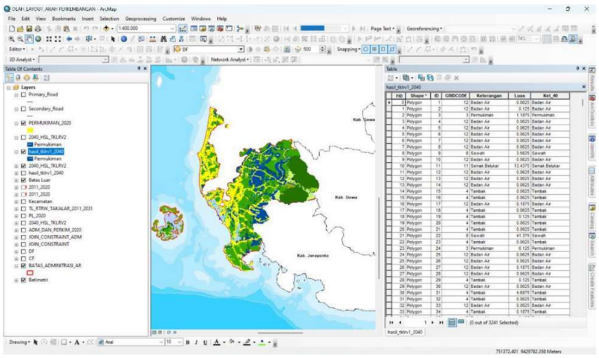
Sebagai landasan melakukan penelitian yang akan dilakukan maka dibahas teori yang berkaitan. Berikut ini kami uraikan landasan teori dari karya ini.

2.2.1 Data Geospasial

Data geospasial merupakan representasi digital dari objek, fenomena, atau kejadian yang memiliki referensi lokasi di permukaan bumi dan dinyatakan dalam sistem koordinat tertentu, sehingga informasi yang disajikan tidak hanya menjelaskan "apa" objeknya, tetapi juga "di mana" objek tersebut berada secara keruangan. Konsep geospasial menekankan aspek lokasi, letak, dan posisi objek yang berada di bawah, pada, atau di atas permukaan bumi, yang dinyatakan melalui georeference atau referensi ruang kebumian, sehingga data dapat dipakai untuk kebutuhan analisis berbasis ruang. Data geospasial umumnya tersusun atas dua komponen utama, yaitu informasi lokasi (spasial) yang terkait dengan koordinat (misalnya lintang–bujur atau koordinat XYZ) dan informasi atribut (deskriptif/non-spasial) yang menjelaskan karakteristik objek, seperti jenis, nama, kode, luas, atau informasi identifikasi lain yang melekat pada lokasi tersebut. Dalam konteks dokumen teknis geospasial,

komponen atribut sering memuat istilah dan kode khusus yang menjadi kunci pencarian. Dari sisi bentuk penyajian, data geospasial dapat disajikan dalam format vektor (menggunakan titik, garis, dan bidang/poligon) maupun raster (merepresentasikan objek geografis dalam bentuk piksel), di mana perbedaan format ini memengaruhi cara data direpresentasikan dan dimanfaatkan (Moh. Erkamim et al., 2024).

Informasi Geospasial (IG) adalah data geospasial yang telah diolah sehingga dapat digunakan sebagai dasar perumusan kebijakan, pengambilan keputusan, dan pelaksanaan kegiatan yang berkaitan dengan ruang kebumian; penyelenggaraan IG dibagi menjadi Informasi Geospasial Dasar (IGD) yang memuat objek fisik dasar yang relatif stabil dalam jangka waktu tertentu, dan Informasi Geospasial Tematik (IGT) yang menyajikan tema tertentu berdasarkan IGD. Uraian ini relevan dengan penelitian karena dokumen SK Geospasial memuat standar dan struktur informasi yang banyak berisi entitas atribut dan kode teknis, sehingga pemahaman tentang komponen data geospasial (spasial–atribut) serta klasifikasi IGD/IGT menjadi landasan untuk merancang sistem informasi pencarian yang mampu mengekstraksi dan menelusuri entitas penting secara lebih terarah sesuai struktur dokumen (Pambudi, 2022). Berikut adalah gambar nya



Gambar 2. 1 Visualisasi Data Geospasial (Moh. Erkamim et al., 2024)

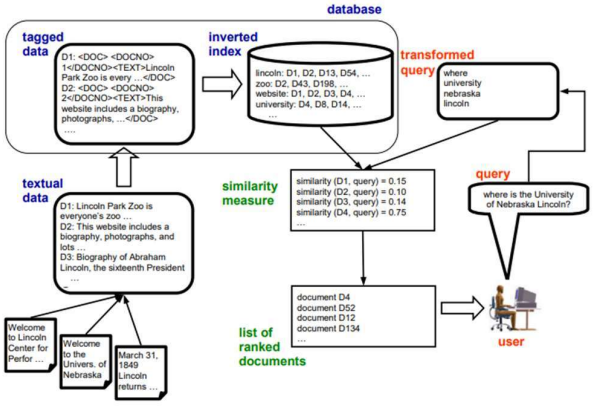
2.2.2 Information Retrieval (IR)

Information Retrieval (IR) atau temu balik informasi merupakan bidang ilmu yang mempelajari proses pencarian dan pengambilan dokumen atau informasi yang relevan dari suatu kumpulan dokumen berdasarkan kebutuhan pengguna. Tujuan utama dari Information Retrieval adalah membantu pengguna menemukan informasi yang paling sesuai dari sekumpulan data yang besar dan terstruktur maupun tidak terstruktur, dengan cara memproses kueri yang diberikan dan mencocokkannya dengan representasi dokumen yang tersimpan dalam system. Secara konseptual, sistem Information Retrieval terdiri atas beberapa komponen utama, yaitu dokumen, kueri, indeks, dan mekanisme pencocokan (matching). Dokumen merupakan sumber informasi yang disimpan dalam sistem, sedangkan kueri adalah representasi kebutuhan informasi pengguna dalam bentuk kata atau frasa. Indeks digunakan untuk merepresentasikan dokumen dalam bentuk struktur data tertentu agar proses pencarian dapat dilakukan secara efisien. Selanjutnya, mekanisme matching berfungsi untuk menghitung tingkat kecocokan antara kueri dan dokumen, yang kemudian digunakan sebagai dasar penentuan relevansi hasil pencarian (Arifin et al., 2022)

Proses Information Retrieval secara umum meliputi tahap pemrosesan dokumen, pemrosesan kueri, pencocokan, dan penyajian hasil. Pada tahap pemrosesan dokumen, isi dokumen dianalisis dan direpresentasikan ke dalam bentuk indeks agar mudah ditelusuri. Tahap pemrosesan kueri bertujuan menyesuaikan masukan pengguna dengan representasi dokumen yang ada. Tahap pencocokan merupakan inti dari sistem IR, karena pada tahap inilah sistem menilai kesesuaian antara kueri dan dokumen, sebelum akhirnya hasil ditampilkan kepada pengguna berdasarkan tingkat relevansi tertentu (Arifin et al., 2022; Hambarde et al., 2023). Dalam praktiknya, pendekatan pencocokan dalam Information Retrieval dapat dibedakan menjadi exact matching dan inexact matching. Exact matching mensyaratkan kesamaan karakter atau kata secara persis antara kueri dan dokumen, sehingga metode ini bekerja optimal ketika penulisan kueri sesuai dengan teks dokumen. Namun, pendekatan ini memiliki keterbatasan ketika pengguna melakukan kesalahan pengetikan, menggunakan singkatan, atau variasi istilah yang berbeda.

Oleh karena itu, banyak penelitian IR modern mulai menekankan pentingnya pendekatan pencocokan yang lebih fleksibel dan toleran terhadap variasi input pengguna (Kurgan, 2007).

Pada dokumen teknis dan administratif yang bersifat panjang serta kaya akan istilah khusus, seperti dokumen kebijakan atau dokumen geospasial, tantangan Information Retrieval menjadi lebih kompleks. Struktur dokumen yang formal dan penggunaan istilah teknis menyebabkan pencarian berbasis kecocokan eksak sering kali gagal memenuhi kebutuhan pengguna. Literatur Information Retrieval modern menegaskan bahwa sistem IR perlu menyesuaikan metode pencocokan dengan karakteristik data dan konteks penggunaan agar hasil pencarian tetap relevan dan praktis (Najork, 2023). Berikut merupakan gambar dari Bagian Sistem Temu Kembali Informasi

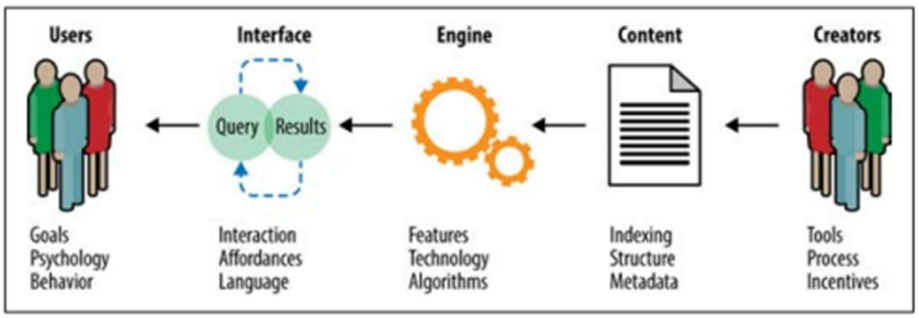


Gambar 2. 2 Bagian Sistem Temu Kembali Informasi (Kurgan, 2007)

2.2.3 Sistem Informasi Pencarian

Sistem informasi pencarian merupakan bagian dari arsitektur informasi yang membantu pengguna berpindah dan menemukan informasi dalam suatu lingkungan dokumen (Rosenfeld et al., 2015). Dalam arsitektur informasi, pengguna dapat menemukan informasi melalui dua jalur utama, yaitu menelusuri melalui sistem navigasi (misalnya menu dan tautan) atau melakukan pencarian melalui sistem pencarian. Sistem navigasi memfasilitasi aktivitas browsing, sedangkan sistem pencarian memfasilitasi aktivitas searching dengan menjalankan kueri terhadap indeks yang sudah disiapkan (Rosenfeld et al., 2015). Kueri (*query*) adalah masukan

yang diberikan pengguna kepada sistem pencarian untuk menyatakan informasi yang ingin ditemukan. Kueri dapat berupa kata, frasa, kode, atau pola tertentu yang mewakili kebutuhan informasi pengguna. Dalam sistem pencarian, kueri digunakan sebagai acuan untuk mengeksekusi pencarian pada indeks yang telah dibangun, sehingga sistem dapat mengembalikan hasil yang paling relevan sesuai masukan pengguna (Rosenfeld et al., 2015).Dalam praktiknya, mesin melakukan pencarian dengan cara mencocokkan pola yang diberikan dengan tabel indeks (teks) yang tersedia, sehingga pengguna tidak perlu lagi melakukan pencarian manual yang cenderung lambat dan rentan terlewat, terutama ketika dokumen yang dihadapi berjumlah banyak atau memiliki struktur panjang dan kompleks. Sejalan dengan itu, pada konteks sistem pencarian modern, pendekatan *text matching* juga digunakan untuk mengukur tingkat kesamaan (similarity) antara teks kueri dan data yang dicari, lalu menampilkan hasil yang melampaui ambang batas tertentu agar keluaran sistem tetap relevan dan mudah dipahami pengguna (Dharma Putra et al., 2025). Cara kerja sistem informasi pencarian secara umum dapat dijelaskan sebagai berikut. Pertama, sistem menyiapkan konten dengan mengekstraksi teks dari dokumen dan membentuk indeks agar informasi dapat ditelusuri secara terstruktur. Kedua, pengguna memasukkan kueri melalui antarmuka. Ketiga, sistem mengeksekusi kueri terhadap indeks dan menghitung tingkat kecocokan antara kueri dan kandidat data. Keempat, sistem menyaring dan menyajikan hasil yang relevan, disertai konteks atau rujukan lokasi sumber agar pengguna dapat langsung menuju bagian dokumen yang memuat informasi tersebut (Rosenfeld et al., 2015).



Gambar 2. 3 Alur Proses Sistem Informasi Pencarian (Rosenfeld et al., 2015)

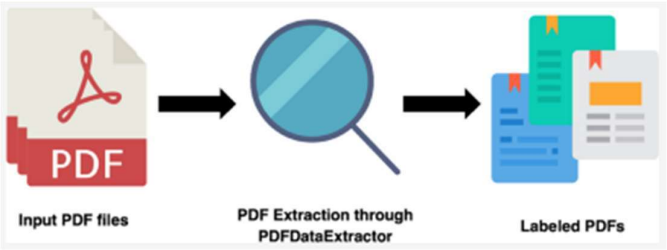
Dalam penelitian ini, sistem informasi pencarian diarahkan untuk dokumen teknis geospasial (misalnya SK) yang memuat banyak istilah, kode, serta entitas spesifik, sehingga dibutuhkan mekanisme pencarian yang tidak hanya bergantung pada kecocokan identik, tetapi mampu mengukur kemiripan antar string agar hasil pencarian tetap muncul meskipun terdapat variasi penulisan. Salah satu pendekatan yang relevan adalah FSM, karena FSM merupakan metode pencarian string dengan proses pendekatan terhadap pola string yang dicari, sehingga dapat membantu menghasilkan keluaran pencarian yang lebih valid dan mudah dipahami dibanding pencarian yang sepenuhnya kaku. Selain itu, penggunaan FSM juga dilaporkan memudahkan proses pencarian keyword/pattern pada sistem berbasis teks (meskipun studi tertentu membahasnya pada konteks layanan tanya-jawab), sehingga secara konsep mendukung kebutuhan navigasi informasi dokumen yang menuntut pencarian cepat dan toleran terhadap variasi input (Mulyatun et al., 2021).

2.2.4 Ekstraksi Teks dari Dokumen PDF

Ekstraksi teks dari dokumen PDF dipelajari dalam konteks arsitektur Local Retrieval-Augmented Generation (RAG), yang dirancang untuk membangun Ekstraksi teks dari dokumen PDF merupakan tahapan awal yang penting dalam sistem pengolahan dan pencarian informasi berbasis dokumen digital. Dokumen PDF pada umumnya bersifat read-only dan tidak memiliki struktur data eksplisit seperti basis data, sehingga informasi di dalamnya perlu diekstraksi terlebih dahulu agar dapat diproses secara komputasional. Pada dokumen resmi seperti Surat Keputusan (SK) Geospasial, struktur dokumen sering kali kompleks, terdiri dari teks naratif, daftar bernomor, serta format multi-baris yang tidak selalu konsisten. Oleh karena itu, proses ekstraksi teks menjadi krusial untuk mengubah konten PDF menjadi data tekstual yang dapat dianalisis dan dicari kembali (Chida et al., 2024).

Proses ekstraksi teks pada penelitian ini dilakukan menggunakan pustaka PyMuPDF (fitz), yang mampu membaca dan mengambil teks langsung dari halaman dokumen PDF secara efisien. PyMuPDF bekerja dengan mengekstraksi konten teks berdasarkan representasi internal dokumen PDF tanpa melakukan proses optical

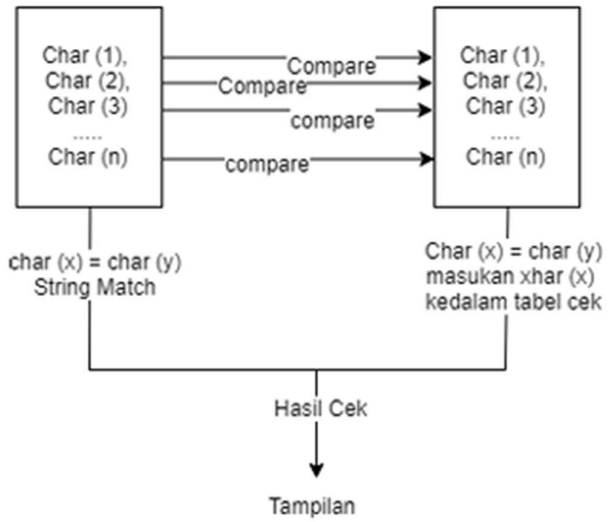
character recognition (OCR), sehingga sesuai untuk dokumen digital resmi yang teksnya masih dapat dibaca secara langsung. Pendekatan ini sejalan dengan penelitian yang membahas pemrosesan dokumen PDF terstruktur, di mana teks diekstraksi terlebih dahulu sebelum dilakukan pemrosesan lanjutan berbasis pola (pattern-based processing). Setelah teks berhasil diekstraksi dari dokumen PDF, tahap selanjutnya adalah pemrosesan teks berbasis pola menggunakan Regular Expressions (Regex). Regex digunakan untuk mengidentifikasi dan memisahkan bagian-bagian penting dalam dokumen, seperti judul bidang, nama produsen data, serta daftar Informasi Geospasial Tematik (IGT). Teknik ini umum digunakan pada dokumen resmi yang memiliki pola penulisan relatif konsisten namun tidak sepenuhnya terstruktur. Dengan Regex, teks hasil ekstraksi dapat disaring, dibersihkan dari noise, serta dipetakan ke dalam struktur tabel yang lebih terorganisasi. Pendekatan ekstraksi teks berbasis PyMuPDF dan Regex ini memungkinkan sistem untuk membangun representasi data terstruktur dari dokumen PDF SK Geospasial, yang selanjutnya digunakan sebagai basis pencarian informasi. Data hasil ekstraksi kemudian dapat diintegrasikan dengan metode pencocokan string seperti FSM untuk mendukung pencarian meskipun terdapat variasi penulisan atau kesalahan ketik pada masukan pengguna. Dengan demikian, proses ekstraksi teks berperan sebagai fondasi utama dalam membangun sistem pencarian dokumen geospasial yang efektif dan responsif terhadap kebutuhan pengguna (Zhu et al., 2022). Berikut merupakan gambar Proses ekstraksi data dari dokumen PDF



Gambar 2. 4 Proses ekstraksi data dari dokumen PDF (Zhu et al., 2022)

2.2.5 Fuzzy String Matching (FSM)

Fuzzy String Matching (FSM) merupakan metode pencarian string yang menggunakan pendekatan kemiripan terhadap pola string yang ingin dicari, sehingga termasuk ke dalam kategori inexact matching. Metode ini memungkinkan sistem untuk tetap menemukan string yang relevan meskipun susunan karakter antara input pengguna dan data yang tersimpan di dalam basis data tidak sepenuhnya sama. Oleh karena itu, FSM sangat berguna dalam mengatasi kesalahan penulisan (typo), variasi ejaan, maupun transposisi karakter yang umum terjadi pada input pengguna. Kunci utama dalam FSM adalah penggunaan fungsi kesamaan (similarity function) yang berfungsi untuk mengukur tingkat kemiripan antara string masukan dan string target yang tersimpan di dalam kamus atau basis data. Nilai kemiripan yang dihasilkan kemudian digunakan sebagai dasar dalam menentukan apakah suatu string dapat dianggap cukup mendekati untuk ditampilkan sebagai hasil pencarian. Pendekatan ini bersifat aproksimatif, sehingga sistem tidak bergantung pada kecocokan karakter secara eksak. Secara konseptual, FSM sejalan dengan prinsip Logika Fuzzy yang memperkenalkan gagasan kebenaran parsial, di mana suatu kondisi tidak hanya dinilai benar atau salah, tetapi memiliki tingkat kedekatan tertentu. Dalam konteks pencocokan string, konsep ini diwujudkan dalam bentuk nilai kemiripan yang berada pada rentang tertentu. Untuk menghitung tingkat kemiripan tersebut, berbagai algoritma dapat digunakan, salah satunya adalah Jarak Levenshtein (Edit Distance), yang mengukur jumlah operasi minimum yang diperlukan untuk mengubah satu string menjadi string lain, dan sering dijadikan dasar dalam perhitungan kemiripan string. (Dharma Putra et al., 2025).



Gambar 2. 5 Alur Pencocokan String (Noor Fauzy, 2019)

2.2.5.1 Prinsip Kemiripan String

Metode FSM ialah metode pencarian string yang memakai pendekatan terhadap suatu pola dari string yang ingin dicari. Melakukan pencarian terhadap string yang sama dan atau string lain yang mendekati dalam database. Kunci dari metode ini ialah dengan memutuskan bahwa sebuah string yang ingin dicari mempunyai persamaan terhadap string yang tersimpan dalam sebuah penampung atau database, walaupun tidak sama persis pola karakternya. untuk menentukan dan memutuskan tingkat kesamaan dari sebuah string, maka diperlukan sebuah "Fungsi Persemaan" (Similarity function) dimana fungsi inilah yang akan bertugas untuk menentukan dan memutuskan string akhir dari sebuah pencarian string melalui hasil pendekatan (Aproksimasi) (Henri Saputro et al., 2021). Cara kerja dari fuzzy ini ialah mencocokan kemiripan kata dari pattern dan text yang di masukkan. Tahap pencocokannya sebagai berikut:

1. String dicocokan dengan kamus kata yang ada dalam database dan jika memlikiki kesamaan dan tidak ada perbedaan maka akan ditampilkan

Contoh : Kamus = KULIAH

Kata yang dicari = KULIAH

0	1	2	3	4	5
K	U	L	I	A	H
K	U	L	I	A	H

Gambar 2. 6 Analisa FSM (Henri Saputro et al., 2021)

2. Apabila belum ditemukan maka di cek kesamaan huruf awal, huruf akhir dan memeiliki panjang yang sama, jika ditemukan maka ditampilkan.

Contoh : Kamus = KULIAH

Kata yang dicari = KULIYA

0	1	2	3	4	5
K	U	L	I	A	H
K	U	L	I	Y	A

0	1	2	3	4	5
K	U	L	I	A	H
K	U	L	I	Y	A

0	1	2	3	4	5
K	U	L	I	A	H
K	U	L	I	Y	A

0	1	2	3	4	5
K	U	L	I	A	H
K	U	L	I	Y	A

0	1	2	3	4	5
K	U	L	I	A	H
K	U	L	I	Y	A

0	1	2	3	4	5
K	U	L	I	A	H
K	U	L	I	Y	A

Gambar 2. 7 Analisa FSM (Henri Saputro et al., 2021)

Untuk pencocokan kata diatas dimulai dari index ke-0, jika sama maka dilanjutkan ke i+1 yaitu index ke -2, hingga index ke -3 ditemukan kesamaan

dan pada index ke 4 dan ke 5 tidak sama, jika presentase kata yang cocok/sama lebih dari 50% maka ditampilkan.

3. Apabila belum ditemukan maka di cek huruf awal dan akhir dengan menhiraukan panjang huruf, apabila ditemukan maka ditampilkan.

Contoh : Kamus = NAIK PANGKAT

Kata yang dicari = PANGKAT

Dilakukan pengecekan untuk index ke-0 dari Kamus dan kata yang di cari dan tidak ditemukan kesamaan maka berlanjut ke index selanjutnya.

0	1	2	3	4	5	6	7	8	9	10	11
N	A	I	K		P	A	N	G	K	A	T
P	A	N	G	K	A	T					

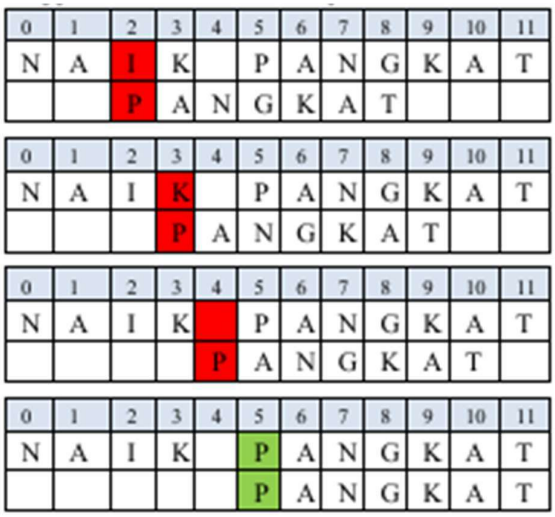
Gambar 2. 8 Analisa FSM (Henri Saputro et al., 2021)

Pada index ke-2 juga masih belum ditemukan kesamaan, lanjut ke index ke-3 dan seterusnya hingga ditemukan kata cocok pertama.

0	1	2	3	4	5	6	7	8	9	10	11
N	A	I	K		P	A	N	G	K	A	T
	P	A	N	G	K	A	T				

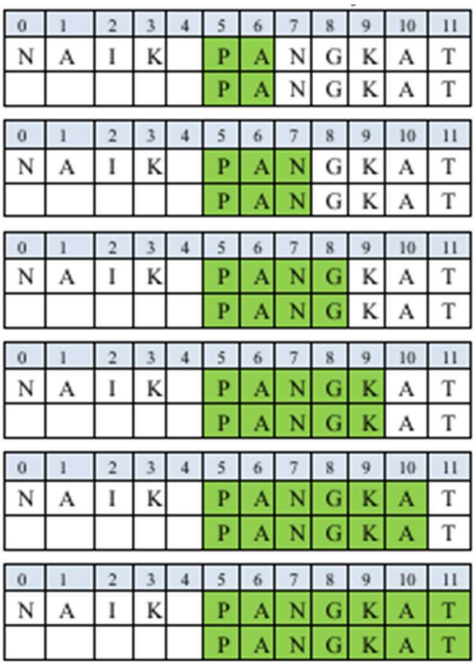
Gambar 2. 9 Analisa FSM (Henri Saputro et al., 2021)

Pada index ke-5 ditemukan kata pertama yang cocok kemudian dicek index berikutnya.



Gambar 2. 10 Analisa FSM (Henri Saputro et al., 2021)

Dan ternyata pada index ke-5 sampai index ke 11 ditemukan kata yang mirip maka kata ditampilkan.



Gambar 2. 11 Analisa FSM (Henri Saputro et al., 2021)

4. Apabila masih belum ditemukan atau tidak ada kemiripan dengan kamus data maka ditampilkan seuai dengan kata ayng dimasukkan pengguna. Dalam aplikasi ini akan dimunculkan pesan “surat tidak ditemukan”.

2.2.5.2 Levenshtein Distance

Levenshtein Distance, juga dikenal sebagai edit distance, adalah sebuah teknik yang digunakan untuk mengukur seberapa besar perbedaan antara dua string. Konsep dasar dari Levenshtein Distance adalah menghitung jumlah operasi yang diperlukan untuk mengubah satu string menjadi string lain dengan tiga jenis operasi yang diperbolehkan: penyisipan karakter, penghapusan karakter, dan penggantian karakter. Semakin kecil nilai Levenshtein Distance antara dua string, semakin mirip kedua string tersebut. Pada dasarnya, Levenshtein Distance berfungsi untuk memperkirakan seberapa banyak perubahan yang perlu dilakukan pada satu string agar menjadi string lain, yang merupakan dasar penting dalam aplikasi FSM. Teknik ini sangat berguna dalam situasi di mana pencocokan yang tepat tidak memungkinkan, seperti saat terjadi kesalahan pengetikan (typo) atau variasi ejaan. Sebagai contoh, dalam pencarian dokumen, pengguna mungkin mengetikkan kata dengan kesalahan kecil, namun sistem masih dapat menemukan kata yang dimaksud dengan membandingkan jarak Levenshtein antara kata yang dimasukkan dan kata yang ada di dalam database (Muhsin et al., 2025).

Dalam penerapannya, algoritma Levenshtein mengukur perbedaan antara dua kata atau string dengan menghitung jumlah minimum langkah yang diperlukan untuk mengubah satu kata menjadi kata lainnya. Misalnya, dalam sistem pencarian berbasis teks, jika pengguna mencari kata yang memiliki kesalahan penulisan, Levenshtein Distance memungkinkan sistem untuk memperkirakan kemiripan antara kata yang dimasukkan dengan kata yang ada di database meskipun ada kesalahan kecil dalam ejaan. Levenshtein Distance sangat mendukung dalam pencocokan string yang lebih fleksibel,

terutama dalam konteks sistem yang memerlukan pencarian teks toleran terhadap kesalahan ketik atau variasi penulisan. Dalam penelitian ini, Levenshtein Distance digunakan untuk mengukur kedekatan antara entitas dalam dokumen geospasial, memungkinkan sistem untuk memberikan hasil pencarian yang lebih relevan meskipun ada variasi penulisan atau kesalahan ketik (Dharma Putra et al., 2025).

2.2.5.3 Perhitungan Tingkat Kemiripan String

Prinsip utama dalam FSM adalah pengukuran tingkat kemiripan (similarity) antara string masukan pengguna dengan string yang tersimpan di dalam basis data. Tingkat kemiripan ini tidak ditentukan berdasarkan kecocokan karakter secara mutlak, melainkan melalui pendekatan kuantitatif yang menghitung seberapa besar perbedaan antarstring masih dapat ditoleransi sebagai satu kesamaan konteks. Oleh karena itu, FSM memerlukan suatu fungsi kesamaan (similarity function) sebagai dasar pengambilan keputusan dalam proses pencarian. Salah satu pendekatan yang banyak digunakan untuk mengukur kemiripan string adalah Jarak Levenshtein (Levenshtein Distance), yaitu ukuran jumlah operasi minimum yang diperlukan untuk mengubah satu string menjadi string lain. Operasi tersebut meliputi penyisipan (insertion), penghapusan (deletion), dan penggantian (substitution) karakter. Semakin kecil nilai jarak Levenshtein antara dua string, maka semakin tinggi tingkat kemiripan di antara keduanya (Putra et al., 2025).

Untuk memperoleh nilai kemiripan dalam bentuk persentase agar mudah diinterpretasikan, nilai jarak Levenshtein dinormalisasi terhadap panjang maksimum string yang dibandingkan. Berdasarkan penelitian (Muhsin et al., 2025), tingkat kemiripan string dapat dihitung menggunakan persamaan sebagai berikut:

$$B = 1 - \frac{d[m,n]}{max(|S|,|T|)} \times 100\% \quad (2.1)$$

dengan keterangan:

1. B adalah nilai tingkat kemiripan (*similarity score*) dalam persen,
2. $d[m, n]$ adalah jarak Levenshtein antara string masukan S dan string pembanding T,
3. $|S|$ dan $|T|$ masing-masing menyatakan panjang string masukan dan string pembanding.

Nilai kemiripan sebesar 100% menunjukkan bahwa kedua string identik, sedangkan nilai yang lebih rendah mengindikasikan adanya perbedaan karakter baik dari segi jumlah maupun susunan. Dengan mekanisme ini, sistem berbasis FSM tetap mampu mengembalikan hasil pencarian yang relevan meskipun terjadi kesalahan penulisan (*typo*), variasi ejaan, atau perbedaan panjang string antara kueri pengguna dan data yang tersimpan.

Untuk memperoleh nilai persentase perubahan antara dua string, nilai AE dihitung berdasarkan jumlah bit yang berubah dibandingkan dengan jumlah bit total dari kedua string yang dibandingkan. Berdasarkan rumus yang digunakan dalam penelitian ini, tingkat perubahan antara dua string dapat dihitung menggunakan persamaan sebagai berikut:

$$AE = \frac{\text{Jumlah bit yang berubah}}{\text{Jumlah bit total}} \times 100\% \qquad (2.2)$$

Dengan keterangan:

a. AE: Average Error (tingkat kesalahan rata-rata) berdasarkan perbandingan bit atau karakter.

b. Jumlah bit yang berubah adalah jumlah karakter yang berbeda antara string yang dibandingkan,

c. Jumlah bit total adalah jumlah total karakter yang ada pada kedua string yang dibandingkan.

Nilai AE sebesar 0% menunjukkan bahwa kedua string identik, sementara nilai yang lebih tinggi menunjukkan adanya perbedaan karakter antara kedua string tersebut. Semakin besar nilai AE, semakin besar perbedaan antara kedua string. Dengan mekanisme ini, sistem dapat menilai seberapa besar perbedaan yang ada antara dua string, baik dalam hal kesalahan penulisan, variasi karakter, atau perbedaan lainnya.

2.2.5.4 Toleransi Kesalahan Penulisan

Toleransi kesalahan penulisan (error tolerance) merupakan kemampuan sistem pencarian untuk tetap memberikan hasil yang relevan meskipun terdapat perbedaan antara kata kunci yang dimasukkan oleh pengguna dengan data yang tersimpan di dalam basis data. Kesalahan penulisan tersebut dapat berupa typo, penghilangan atau penambahan karakter, pertukaran urutan huruf (transposition), serta variasi ejaan yang umum terjadi dalam penggunaan istilah teknis. Pada sistem pencarian berbasis exact matching, perbedaan sekecil apa pun pada susunan karakter dapat menyebabkan kegagalan pencarian karena sistem hanya mengenali kecocokan string yang identik. Penggunaan FSM dalam konteks dokumen geospasial sangat relevan karena dokumen-dokumen ini sering berisi entitas teknis yang spesifik, seperti kode lokasi, nama produsen data, dan kode atribut, yang cenderung memiliki variasi penulisan. Dengan menggunakan FSM, sistem informasi pencarian dapat melakukan pencarian yang lebih efektif dan fleksibel, bahkan ketika pengguna memasukkan variasi istilah atau mengalami kesalahan pengetikan. Dengan demikian, FSM sangat penting dalam meningkatkan efektivitas temu balik informasi, mempercepat pencarian data yang relevan, dan mengurangi frustrasi pengguna akibat kesalahan penulisan (Henri Saputro et al., 2021; Putra et al., 2025).

2.2.6 Framework Aplikasi Web (Streamlit)

Streamlit adalah *framework* sumber terbuka (*open source*) berbasis Python yang secara spesifik dirancang untuk memfasilitasi pembuatan aplikasi *web* interaktif dal-

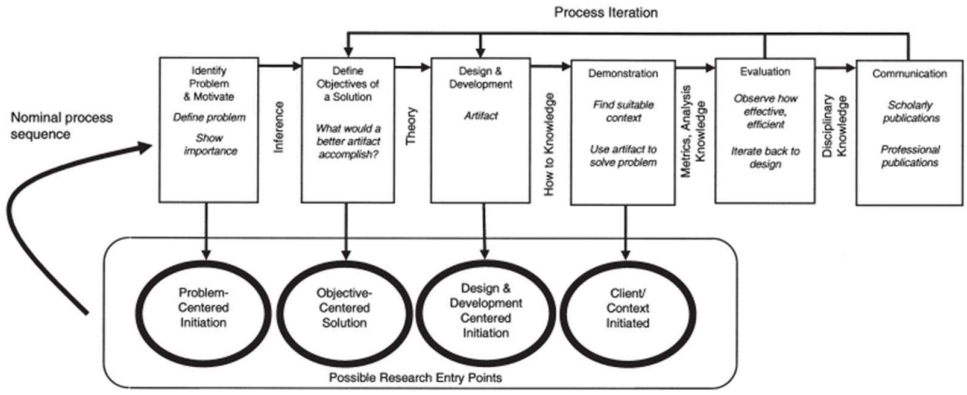
am bidang *data science* dan *machine learning*. Keunggulan utama Streamlit terletak pada kemudahan penggunaannya, karena memungkinkan pengguna yang hanya menguasai dasar-dasar Python untuk membangun antarmuka visual, tanpa memerlukan keahlian teknis dalam pengembangan *front-end* seperti HTML, CSS, atau JavaScript. Framework ini mendukung integrasi yang mulus dengan pustaka Python populer, seperti *pandas* dan *Matplotlib*, untuk mendukung pengolahan data dan visualisasi. Streamlit banyak diterapkan dalam berbagai konteks, termasuk perancangan *website* untuk prediksi (misalnya prediksi jumlah pengeluaran mahasiswa Generasi Z), serta untuk mengimplementasikan hasil analisis *data mining* (seperti algoritma Apriori) secara *real-time* dengan visualisasi interaktif. Aplikasi yang dikembangkan dengan Streamlit dapat *di-deploy* melalui layanan *Streamlit Cloud*, sehingga mudah diakses oleh *stakeholder* tanpa perlu instalasi lokal yang rumit (Devalio et al., 2025).

2.2.7 Blackbox Testing

Blackbox Testing (BBT) adalah metode pengujian perangkat lunak yang berfokus pada spesifikasi fungsional dari aplikasi tanpa melihat struktur internal atau kode program. Pengujian ini juga dikenal sebagai *functional testing* atau *specification-based testing*, dan bertujuan untuk memastikan bahwa sistem berfungsi sesuai dengan spesifikasi yang telah ditetapkan dan memenuhi kebutuhan fungsional pengguna. BBT menilai kinerja sistem dari sudut pandang pengguna akhir.Teknik dan Penerapan: Teknik yang umum digunakan dalam BBT adalah *Equivalence Partitioning*, yang membagi input menjadi beberapa partisi data yang ekuivalen untuk mengidentifikasi skenario pengujian. Keberhasilan pengujian BBT sangat menentukan tingkat kelayakan aplikasi sebelum diluncurkan. BBT diterapkan untuk menguji fungsionalitas fitur-fitur utama sistem, seperti proses *login*, registrasi, pemberian rekomendasi laptop, dan fitur-fitur pada sistem peminjaman kendaraan dan e-PAS bandara. Pada pengujian *chatbot* rekomendasi wisata, BBT digunakan untuk menguji akurasi respons, yang pada satu penelitian mencapai 100% akurasi (Shadiq et al., 2021).

2.2.8 Design Science Research Methodology (DSRM)

Design Science Research Methodology (DSRM) adalah sebuah metodologi yang berorientasi desain sistem informasi. DSRM juga merupakan kerangka prosedur yang digunakan untuk mempermudah penelitian di bidang teknologi informasi yang digunakan sebagai proses pemahaman serta mengulas untuk mengenali dan mengevaluasi hasil penelitian DSRM menurut (Peffers et al., 2007) adalah sebuah metodelogi penelitian di bidang ilmu desain untuk sistem informasi yang menggabungkan antara prinsip, praktik, dan prosedur. Ada beberapa Langkah penelitian nya yaitu: identifikasi masalah dan motivasi (problem identification and motivation), definisi tujuan untuk solusi (objectives of a solution), desain dan pengembangan (design&development), demontrasi (demonstration), evaluasi (evalu ation), dan komunikasi (communication). Tahapan- tahapan tersebut diilustrasikan pada gambar dibawah (Orisa et al., 2023).



Gambar 2. 12 Tahapan Metodologi Penelitian (Orisa et al., 2023)

2.3 Kerangka Pemikiran

Berikut adalah kerangka pemikiran yang disusun untuk menggambarkan langkah-langkah proses analisis data dari awal hingga akhir, dengan tujuan untuk mendapatkan kesimpulan dari data yang sedang dianalisis. Terdapat empat komponen yang menyusun kerangka pemikiran penelitian yaitu Indikator (Indicators), Metode yang diusulkan

(Proposed Method), Tujuan (Objective), dan Pengukuran (Measurement) (Siswipraptini et al., 2023). Pada penelitian ini, langkah yang diimplementasikan ke dalam sistem digambarkan melalui perancangan sistem sebagai berikut :

Berikut ini disajikan kerangka pemikiran untuk penelitian ini. Setiap aliran menunjukkan hubungan yang menarik antara komponen-komponen yang berbeda. Selanjutnya, penjelasan untuk masing-masing komponen pada Gambar 2.9 akan diberikan di bawah ini.

1. Indikator

Metrik merujuk pada parameter yang diuji dalam sebuah penelitian, sementara indikator didefinisikan sebagai variabel input utama yang akan melalui serangkaian proses pengolahan lebih lanjut. Dalam konteks penelitian ini, indikator yang digunakan adalah dokumen Surat Keputusan (SK) resmi dari Kementerian Kehutanan, khususnya SK Nomor 398 Tahun 2024 tentang Standar Data Geospasial dan IGT, SK Nomor 399 Tahun 2024 tentang Penyebarluasan IGT, serta SK Nomor 400 Tahun 2024 tentang Perubahan Produsen Data. Dokumen-dokumen tersebut merupakan acuan teknis nasional yang padat informasi, mencakup entitas seperti kode unsur geospasial (*LCODE*), nama produsen data, struktur atribut, hak akses field, aturan topologi, dan skala peta. Dataset ini menjadi dasar bagi sistem informasi pencarian, karena kompleksitas dan ketebalannya menyebabkan kesulitan dalam pencarian manual oleh staf teknis, peneliti, maupun mahasiswa yang membutuhkan informasi spesifik secara cepat.

2. Proposed method

Metode yang diusulkan mencakup pendekatan yang dipilih berdasarkan kebutuhan terhadap efisiensi, akurasi, dan toleransi terhadap kesalahan manusia. Dalam kerangka penelitian ini, penulis menerapkan metode *FSM* sebagai solusi utama untuk mengatasi permasalahan pencocokan string eksak yang rentan gagal saat terjadi typo atau variasi ejaan. Metode ini sangat relevan karena tidak bergantung pada kesamaan kata persis, melainkan pada tingkat kemiripan antara dua teks, sehingga tetap dapat mencocokkan “C1O18AR” dengan “C1018AR” atau “Produsen Data C1018AR” meskipun ada perbedaan karakter. Proses ini didukung oleh ekstraksi teks otomatis dari file PDF menggunakan library PyMuPDF, parsing entitas kunci, dan integrasi antarmuka

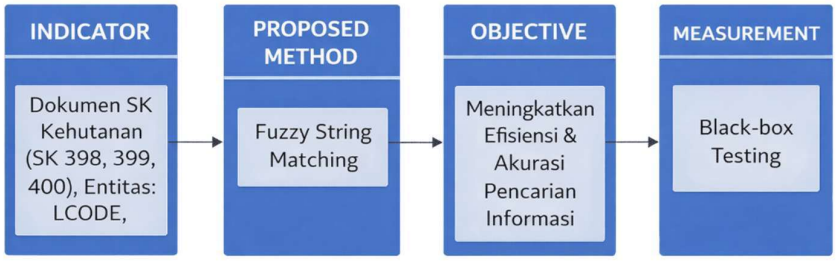
chatbot berbasis web. Pendekatan *FSM* dipilih karena lebih ringan dan efisien dibanding model Natural Language Processing (NLP) berbasis deep learning, terutama karena jumlah dokumen sangat terbatas (hanya 3 buah SK), namun struktur datanya sangat baku dan terstruktur.

3. Objectives

Tujuan penelitian merujuk pada hasil yang ingin dicapai setelah seluruh proses pengolahan dan implementasi selesai. Tujuan yang ingin dicapai dalam penelitian ini adalah mengembangkan prototipe sistem informasi pencarian berbasis chatbot yang mampu mengekstrak, mencari, dan menampilkan entitas penting seperti *LCODE*, produsen data, dan field atribut dari dokumen SK Kementerian Kehutanan secara cepat, akurat, dan interaktif. Sistem ini dirancang agar pengguna dapat mengajukan pertanyaan dalam bahasa alami, misalnya "Siapa produsen data C1O18AR?" atau "Field apa saja di AMDAL?", dan sistem akan memberikan jawaban tanpa harus membaca ratusan halaman PDF secara manual. Dengan demikian, penelitian ini bertujuan untuk meningkatkan efisiensi akses informasi dan mendukung transformasi digital layanan administratif di lingkungan Kementerian Kehutanan.

4. Measurement

Pengukuran merujuk pada langkah evaluasi hasil penelitian yang dilakukan secara objektif. Setelah melewati tahapan seperti ekstraksi data dari dokumen PDF, parsing entitas, implementasi *FSM*, dan pembentukan antarmuka chatbot, tahap validasi sistem menjadi esensial. Pengujian dilakukan menggunakan metode *blackbox testing*, yang fokus pada input-output tanpa melihat struktur internal kode. Metrik evaluasi yang digunakan meliputi: akurasi pencocokan, yaitu seberapa tepat sistem mengidentifikasi entitas yang dimaksud; waktu respons, yaitu waktu rata-rata sistem dalam memberikan jawaban; serta usability, yang dinilai melalui observasi langsung atau kuesioner singkat terhadap pengguna simulasi. Hasil pengujian kemudian dianalisis untuk menilai keberhasilan sistem dalam menjawab kebutuhan pengguna. Evaluasi ini menunjukkan bahwa pendekatan berbasis *FSM* sangat efektif untuk sistem informasi pencarian dengan entitas terstruktur, sebagaimana dibuktikan oleh penelitian (Henri Saputro et al., 2021) dan (Widianto, 2022) yang mencatat peningkatan akurasi hingga 85–90% dalam sistem pencarian dokumen formal.



Gambar 2. 13 Kerangka pemikiran (Siswipraptini et al., 2023)