

Integrasi Basis Data Properti Menggunakan Metode *Schema Matching* dengan Pendekatan *Linguistic* dan *Constraint*

Property Database Integration Using Schema Matching with Linguistic and Constraint

Muhamamd Hariri*, Bambang Krismono Triwijoyo, Galih Hendro Martono
Universitas Bumigora, Mataram, Indonesia

Informasi Artikel:

Diterima: 7 Februari 2025, Direvisi: 10 Juni 2025, Disetujui: 17 Juni 2025

Abstrak-

Latar Belakang: Perkembangan teknologi yang pesat telah mendorong kemajuan di berbagai sektor, termasuk industri properti di Indonesia. Namun, integrasi data properti di Pulau Lombok masih menghadapi tantangan akibat keragaman penamaan atribut, yang menghambat efisiensi dalam pencarian informasi.

Tujuan: Penelitian ini bertujuan untuk mengintegrasikan empat basis data properti (Saduthama, Salva, SJP, dan Garden View) menggunakan metode *schema matching* berbasis pendekatan *linguistic* dan *constraint*.

Metode: Pendekatan linguistik digunakan untuk mengidentifikasi kesamaan antar atribut meskipun memiliki nama yang berbeda, dengan menerapkan teknik *Bigram* yang terbukti efektif dalam mengenali kemiripan atribut dengan ambang batas 0,7. Sementara itu, pendekatan *constraint* mengevaluasi kesesuaian atribut berdasarkan kriteria tambahan seperti tipe data, panjang atribut, nilai *null*, dan keunikan, sehingga memastikan bahwa atribut yang terintegrasi dapat bekerja secara kompatibel. Proses integrasi mencakup tahap *preprocessing*, generalisasi, dan pencocokan atribut.

Hasil: Hasil evaluasi menunjukkan nilai *precision* (P), *recall* (R), dan *F-measure* sebesar 90%, dengan akurasi rata-rata 84%.

Kesimpulan: Hasil ini melampaui penelitian sebelumnya yang mencapai *precision* 100%, *recall* 60%, dan *F-measure* 75%.

Kata Kunci: Integrasi Basis Data; Linguistic; Properti; *Schema Matching*.

Abstract-

Background: The rapid development of technology has driven progress across various sectors, including the property industry in Indonesia. However, property data integration on Lombok Island still faces challenges due to the diversity of attribute naming, which hinders efficient information retrieval.

Objective: This study aims to integrate four property databases (Saduthama, Salva, SJP, and Garden View) using a schema matching method based on linguistic and constraint approaches.

Methods: The linguistic approach is used to identify similarities between attributes, even when their names differ, using the Bigram technique, which proved effective in identifying attribute similarities with a threshold of 0.7. Meanwhile, the constraint approach evaluates the compatibility of attributes based on additional criteria such as data type, attribute length, null values, and uniqueness, ensuring that the integrated attributes work compatibly. The integration process includes preprocessing, generalization, and attribute matching.

Result: The evaluation results show precision (P), recall (R), and F-measure of 90%, with an average accuracy of 84%.

Conclusion: This result outperforms previous studies that achieved 100% precision, 60% recall, and 75% F-measure.

Keywords: Constraint; Database Integration; Linguistic; Property; Schema Matching.

Penulis Korespondensi:

Muhamamd Hariri,
Universitas Bumigora, Mataram, Indonesia,
Email: haririmuhammad65@gmail.com

How to Cite: M. Hariri, B. K. Triwijoyo, dan G. H. Martono, "Integrasi Basis Data Properti Menggunakan Metode *Schema Matching* dengan Pendekatan *Linguistic* dan *Constraint*," *Jurnal Bumigora Information Technology (BITe)*, vol. 7, no. 1, pp. 37-50, Jun 2025. doi: [10.30812/bite.v7i1.4872](https://doi.org/10.30812/bite.v7i1.4872).

This is an open access article under the CC BY-SA license (<https://creativecommons.org/licenses/by-sa/4.0/>)

1. PENDAHULUAN

Kemajuan pesat di era modern telah memengaruhi berbagai aspek kehidupan manusia, termasuk perkembangan dunia teknologi yang terus berevolusi [1]. Salah satu dampak nyata dari perkembangan ini di Indonesia adalah pendekatan dalam pemasaran properti secara global. Dalam konteks bisnis properti di Indonesia, kemajuan teknologi memberikan dampak positif yang signifikan. Industri properti di Indonesia mengalami pertumbuhan yang pesat, sebagaimana dibuktikan oleh meningkatnya jumlah pembangunan perumahan baru yang ditawarkan dengan harga kompetitif [2]. Berbagai solusi perangkat lunak telah bermunculan untuk memenuhi kebutuhan individu maupun institusi, mendorong industri properti sebagai salah satu pilar utama pembangunan ekonomi di Indonesia [3]. Saat ini, rumah tidak hanya dipandang sebagai tempat tinggal, tempat beristirahat, dan berkumpul bersama keluarga, tetapi juga sebagai aset investasi yang bernilai tinggi [4]. Namun, banyak orang masih menghadapi kesulitan dalam memperoleh informasi perumahan yang andal. Selain itu, metode pemasaran properti secara tradisional, seperti penggunaan baliho dan brosur, dianggap kurang efektif karena menyulitkan calon pembeli dalam memvisualisasikan bentuk rumah yang sebenarnya [5]. Variasi dalam konvensi penamaan atribut properti juga menjadi kendala dalam integrasi dan pencarian informasi yang efisien. Perbedaan ini mencakup elemen seperti nama properti, harga, jumlah kamar mandi, jumlah unit tersedia, dan atribut lainnya [6].

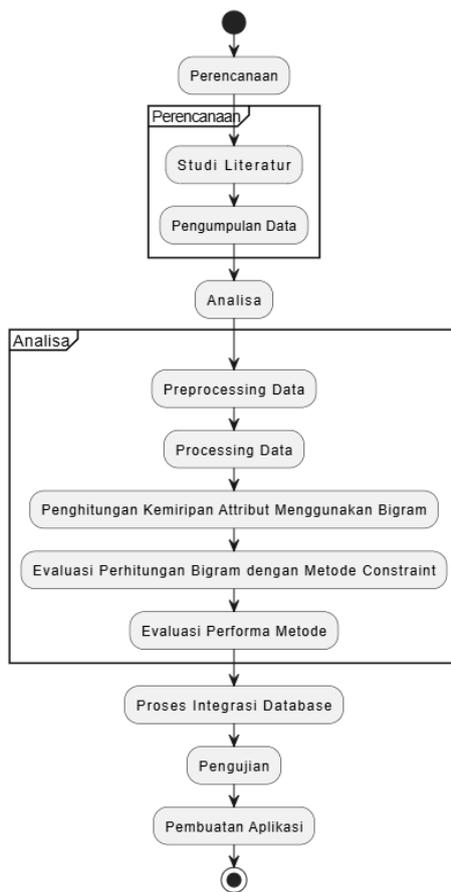
Berdasarkan permasalahan yang telah diuraikan, integrasi basis data properti menjadi langkah penting untuk menghubungkan data properti yang ada, sehingga memudahkan pengguna dalam mencari informasi properti di Lombok. Dalam studi kasus ini, terdapat empat pengembang properti utama, yaitu Sadhutama, Garden View, Salva, dan SJP, yang masing-masing memiliki format dan sistem pencatatan data yang berbeda-beda. Perbedaan ini mencakup penamaan atribut seperti nama properti, harga, luas bangunan, jumlah kamar, hingga status ketersediaan unit. Metode yang digunakan dalam penelitian ini sebagai dasar pengembangan basis data adalah *schema matching*. *Schema matching* adalah teknik yang digunakan untuk mengidentifikasi kesamaan antara atribut atau tabel dalam basis data yang mewakili informasi serupa dari segi bahasa, aturan, atau konten. Metode ini mencakup berbagai pendekatan, seperti linguistik, *constraint*, *instance*, struktur, berbasis aturan, hibrida, dan lainnya. Pendekatan-pendekatan ini dapat diterapkan secara kombinasi atau mandiri [7].

Sejumlah penelitian terkait pengembangan sistem basis data menggunakan metode *schema matching* telah dilakukan dan pembuatan aplikasi properti. Penelitian yang dilakukan oleh [8] pengembangan sistem aplikasi manajemen properti bagian penyewaan unit (Studi Kasus: Pt. Satrand Group Indonesia). Penelitian lainnya [9] yang membuat aplikasi promosi properti berbasis web menggunakan metode waterfall. Penelitian lainnya [10] perancangan dan pengembangan aplikasi jual beli properti berbasis web dengan metode SDLC (Studi Kasus: PT Limitra Indonesia Mega). Penelitian lainnya [11] yang mengembangkan sistem basis data untuk mengintegrasikan tiga basis data dalam sistem informasi manajemen rumah sakit, penelitian ini menggunakan metode linguistik untuk mencocokkan skema antar basis data. Penelitian lainnya [12] mengombinasikan metode berbasis linguistik dan *constraint* dalam *schema matching*, menghasilkan *precision* sebesar 60%, *recall* 100%, dan *F-measure* 75% pada basis data administrasi, akademik, pegawai, penggajian, dan perpustakaan. Penelitian lainnya [13] menerapkan prediktor pencocokan pada dataset *benchmark* berskala besar, tetapi dampak dari prediktor tersebut masih memerlukan evaluasi lebih lanjut. Untuk meningkatkan pemahaman hubungan antar skema, penggunaan teknik *deep learning* dapat meningkatkan akurasi pencocokan skema dengan memungkinkan analisis yang lebih canggih. Penelitian lainnya [14] memperkenalkan kerangka kerja *schema matching* yang dirancang untuk skema heterogen dan kompleks, tetapi hasilnya masih bersifat awal dan belum diuji dalam skenario dunia nyata yang kompleks. Terakhir, penelitian [15] menerapkan metode RESIN-11 pada dataset berita multibahasa, multimedia, dan berbasis peristiwa. Salah satu tantangan utama pendekatan ini adalah ketergantungannya pada kualitas pustaka skema dalam *zero-shot learning*. Untuk mengatasi masalah ini, Penelitian ini menerapkan pendekatan berbasis linguistik dan *constraint* untuk mencocokkan skema data dalam basis data properti yang memiliki atribut berbeda. Dengan metode ini, penelitian bertujuan untuk mempermudah pencarian informasi properti bagi masyarakat, khususnya di Pulau Lombok. Penelitian ini memiliki keunggulan karena tidak

hanya mengkombinasikan kedua metode *schema matching* (linguistik dan *constraint*), tetapi juga mengevaluasi pendekatan ini menggunakan *F-Measure*, menerapkannya dalam domain properti, serta mengimplementasikan hasilnya dalam pengembangan aplikasi.

2. METODE PENELITIAN

Penelitian ini menggunakan metode *schema matching* dengan pendekatan *linguistic* dan *constraint* untuk mengintegrasikan basis data properti. Studi kasus dalam penelitian ini berfokus pada integrasi empat basis data properti, yaitu : Sadhutama, Garden View, Salva, SJP. Metode penelitian dalam penelitian ini dapat dilihat pada Gambar 1. Gambar tersebut dengan jelas menggambarkan beberapa proses penting, termasuk tahap perencanaan awal, analisis mendalam, perancangan sistematis, implementasi terstruktur, dan kesimpulan akhir.



Gambar 1. Metode Penelitian

2.1. Perencanaan

Pada tahap ini, peneliti melakukan dua proses utama: tinjauan literatur dan pengumpulan data. Tinjauan literatur dilakukan untuk memahami konsep *schema matching*, pendekatan *linguistic* dan *constraint* serta bagaimana metode ini telah diterapkan dalam penelitian sebelumnya. Setelah jurnal terkumpul, peneliti melakukan seleksi berdasarkan kriteria tertentu, seperti relevansi dengan penelitian, tahun publikasi, dan tingkat kredibilitas sumber.

2.1.1. Tinjauan Literatur

Pada tahap ini, peneliti melakukan tinjauan literatur terkait *schema matching*, pendekatan linguistik dan *constraint*. Langkah-langkah dalam melakukan tinjauan literatur adalah sebagai berikut:

1. Pencarian Jurnal Referensi

Pencarian jurnal referensi dalam penelitian ini dilakukan berdasarkan penggunaan kata kunci yang relevan dengan topik penelitian, seperti *schema matching*, pendekatan linguistik, *constraint*, dan *instance*. Pencarian jurnal dilakukan melalui platform seperti IEEE dan Google Scholar.

2. Seleksi Jurnal

Pada tahap seleksi jurnal, peneliti mencari jurnal internasional berbahasa Inggris dan jurnal nasional yang diterbitkan dalam lima tahun terakhir. Jurnal yang dipilih harus relevan dengan tema penelitian, yaitu integrasi data properti menggunakan *schema matching* dengan pendekatan linguistik, *constraint*, dan *instance*. Tahap ini dilakukan untuk menilai relevansi jurnal dengan topik penelitian.

2.1.2. Pengumpulan Data

Pada tahap ini, peneliti mengumpulkan data yang akan dianalisis, yaitu basis data properti dari berbagai pengembang, yaitu Salva Property, SJP, Saduthama, dan Garden View. Hasil pengumpulan data berupa brosur dan basis data yang berisi informasi properti. Data yang telah dikumpulkan kemudian diproses untuk digunakan dalam implementasi *schema matching* dengan pendekatan linguistik dan *constraint*.

2.2. Analisis

2.2.1. Praproses Data

Tujuan dari praproses data adalah untuk mengolah data yang telah dikumpulkan pada tahap sebelumnya dengan menghapus atribut yang redundan serta menambahkan atribut baru. Selain itu, dilakukan normalisasi hingga bentuk normal ketiga (*Third Normal Form* atau 3NF). Dari hasil pemodelan konseptual yang telah dibuat, setiap relasi telah memenuhi persyaratan relasional dan memenuhi normalisasi hingga 3NF. Tabel yang telah mencapai 3NF menunjukkan bahwa tidak ada atribut non-kunci yang bergantung pada atribut non-kunci lainnya, sehingga setiap atribut sepenuhnya bergantung pada kunci utama (*primary key*). Selanjutnya, dalam pemodelan logis, setiap tabel yang berelasi telah dihubungkan menggunakan *foreign key* yang merujuk pada *primary key* tabel lain.

2.2.2. Pemrosesan Data

Setelah tahap praproses data selesai, langkah selanjutnya adalah penerapan metode *schema matching* [6]. Proses *schema matching* digunakan untuk menemukan kesamaan di antara berbagai basis data. Pada proses *schema matching* berbasis linguistik, langkah pertama yang dilakukan adalah generalisasi, yaitu dengan memodifikasi nama tabel dan atribut, serta mengubah singkatan menjadi nama lengkap yang lebih standar [16].

2.3. Desain

Pada tahap ini, dilakukan perancangan ulang berdasarkan hasil dari *schema matching*. Proses *schema matching* menghasilkan delapan tabel yang terdiri dari pengguna (*users*), fasilitas (*facilities*), dan properti (*properties*). Tabel-tabel ini kemudian dimodelkan ke dalam *Entity-Relationship Diagram* (ERD). ERD kemudian dianalisis kembali berdasarkan kebutuhan pengguna untuk menambahkan entitas yang mungkin belum ada. Pemodelan konseptual ini dibagi menjadi beberapa subsistem manajemen untuk fokus pada masing-masing basis data: Saduthama, SJP, Salva, dan Garden View. Pemodelan konseptual untuk basis data Saduthama, SJP, Salva, dan Garden View melibatkan transformasi model basis data lama (dari hasil *schema matching*) menjadi model baru. Transformasi ini mencakup penghapusan atribut yang redundan serta penambahan tabel baru.

2.4. Implementasi

Pada tahap ini, rancangan basis data yang telah dikembangkan mulai diimplementasikan. Implementasi dilakukan menggunakan Sistem Manajemen Basis Data (DBMS) MySQL [17]. Pemodelan logis memastikan

bahwa setiap hubungan antar-entitas berbentuk satu ke satu (*one-to-one*) atau satu ke banyak (*one-to-many*), sehingga cocok untuk diterapkan dalam basis data relasional. Selain itu, dilakukan validasi untuk memastikan bahwa basis data memenuhi normalisasi hingga bentuk normal ketiga (3NF). Dari hasil pemodelan konseptual, telah teridentifikasi bahwa setiap hubungan telah memenuhi persyaratan relasional dan mencapai 3NF. Tabel yang mencapai 3NF menunjukkan bahwa tidak ada atribut non-kunci yang bergantung pada atribut non-kunci lainnya, sehingga setiap atribut sepenuhnya bergantung pada kunci utama (*primary key*). Pemodelan logis juga memastikan bahwa setiap tabel yang saling terhubung telah dikaitkan menggunakan *foreign key* yang merujuk pada *primary key* tabel lainnya. Selanjutnya, tahap pemodelan fisik menerjemahkan model logis ini ke dalam tabel yang siap diimplementasikan secara fisik, disesuaikan dengan DBMS yang dipilih (MySQL). Tahap ini mencakup pendefinisian teknis seperti nama tabel, nama kolom, tipe data, panjang data maksimum, nilai *default*, opsi *nullable*, serta atribut tambahan lainnya.

3. HASIL DAN PEMBAHASAN

3.1. Perencanaan

Studi literatur yang dikumpulkan pada penelitian ini adalah terkait dengan *schema matching*, khususnya *schema matching* berbasis *linguistic* dan *constraint*. Selain itu terdapat studi literatur mengenai integrasi basis data. Pada tahap perencanaan ini juga peneliti mengumpulkan data yang digunakan dalam penelitian ini. Adapun data yang digunakan adalah basis data Saduthama, Salva, SJP dan Garden View yang mana jumlah tabel dan atributnya dapat dilihat pada Tabel 1.

Tabel 1. Basis data saduthama, salva, sjp dan garden view

Database	Number of Tables	Number of Attributes
Saduthama	3	21
Salva	3	19
SJP	3	21
Garden View	3	19

3.2. Analisa

3.2.1. Preprocessing data

Preprocessing data bertujuan untuk mengolah data yang telah dikumpulkan pada proses sebelumnya dengan menghilangkan atribut *redundant* dan penambahan atribut baru. Dari hasil pendataan didapatkan beberapa database yang ditunjukkan pada Tabel 1. Semua basis data yang ditemukan masih belum terintegrasi. Basis data pada Tabel 1 kemudian akan memasuki tahap *preprocessing* dan generalisasi. Sebagai contoh, pada Tabel 2, terlihat bahwa nama attribut dari masing-masing basis data mengalami perubahan.

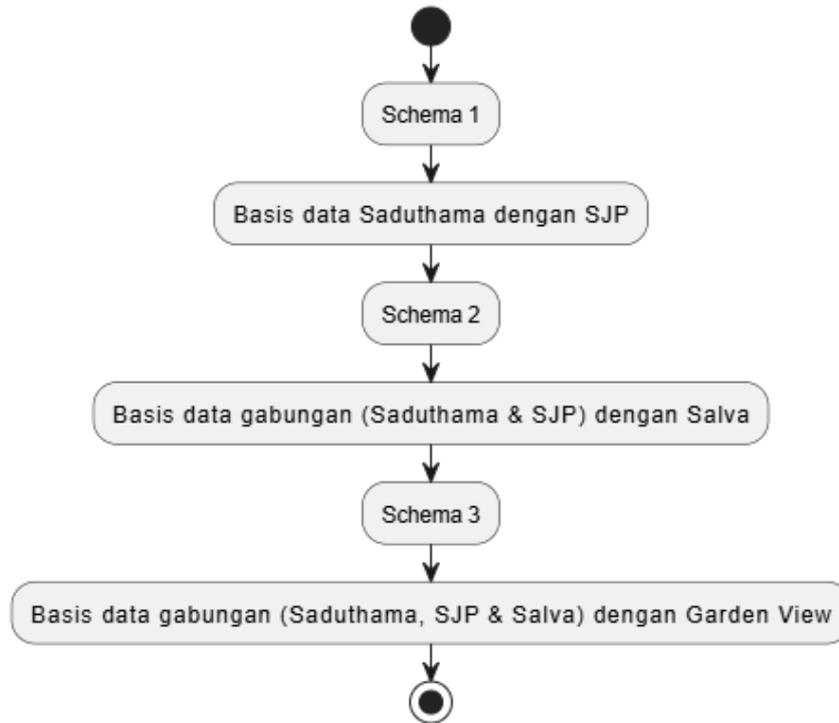
Tabel 2. Proses generalisasi

Basis Data	Nama Tabel	Atribut Sebelumnya	Atribut baru
saduthama	property	Id	id
		Id fasilitas	id fasilitas
		Nama property	nama property
		Developer	developer
		Tipe	tipe
		Harga	harga
		Luas tanah	luas tanah
		Spesifikasi	spesifikasi
		Lokasi	lokasi
		Jenis	jenis property
		Kamar mandi	kamar mandi
		Kamar tidur	kamar tidur
		Ruangan keluarga	ruangan keluarga
Keunggulan	keunggulan		

Basis Data	Nama Tabel	Atribut Sebelumnya	Atribut baru
		Alamat	alamat
salva	property	Id	id
		Id fasilitas	id fasilitas
		Nama Property	nama property
		Developer	developer
		Tipe	tipe
		Harga	harga
		Luas_tanah	luas tanah
		Spesifikasi	spesifikasi
		Lokasi	lokasi
		Jenis	jenis property
		Kamar mandi	kamar mandi
		Kamar tidur	kamar tidur
		Ruangan keluarga	ruangan keluarga
sjp	property	Id	id
		Id fasilitas	id fasilitas
		Name property	name property
		Developer	developer
		Type	type
		Harga	harga
		Luas tanah	luas tanah
		Keterangan	deskripsi
		Lokasi	lokasi
		Jenis	jenis property
		Kamar mandi	kamar mandi
		Kamar tidur	kamar tidur
		Ruangan keluarga	ruangan keluarga
		Spesifikasi	spesifikasi
		Lantai	jumlah lantai
garden view	property	Id	id
		Id fasilitas	id fasilitas
		Nama property	nama property
		Developer	developer
		Type	type
		Harga	harga
		Luas tanah	luas tanah
		Spesifikasi	spesifikasi
		Lokasi	lokasi
		Jenis	jenis property
		Kamar mandi	kamar mandi
		Kamar tidur	kamar tidur
		Ruangan keluarga	ruangan keluarga

3.2.2. Processing data

Setelah dilakukan generalisasi, tahap berikutnya dilakukan *tagging* menggunakan *Bigram* (memecah *string* menjadi dua karakter). Proses ini dilakukan untuk kata yang memiliki makna sama sehingga dianggap sama seperti nama property dengan name property dan lain sebagainya. Penelitian ini mencocokkan empat basis data, sehingga proses *schema matching* dilakukan sebanyak tiga kali. Proses *schema matching* bisa dilihat pada Gambar 2.



Gambar 2. Proses *schema matching*

Tabel 3 menunjukkan proses *bigram* yang dilakukan pada tabel *property* baik dari basis data Saduthama maupun SJP. Hasil pemecahan karakter menunjukkan bahwa nilai kesamaan yang didapatkan dari tabel tersebut adalah 0.9 yang mana nilai kemiripan antar atribut lebih dari *threshold* yaitu 0.7 [11] maka kedua atribut tersebut dapat dianggap sama (duplikasi) sehingga salah satu atribut akan dihapus.

Tabel 3. Contoh Bigram Proses pada Atribut Serupa

Basis data	Tabel	Atribut	Similaritas <i>Bigram</i>
Saduthama	property	nama property	na,am,ma,pr,ro,op,pe,er,rt,ty
SJP	property	name property	na,am,me,pr,ro,op,pe,er,rt,ty
<i>Intersect</i>			9
<i>Similarity</i>			$= \frac{2 \times 9}{10 + 10}$
			$= 0.9$

Tabel 4 hasil pemecahan karakter menunjukkan bahwa nilai kesamaan yang didapatkan dari tabel tersebut adalah 0 yang mana nilai kemiripan antar atribut lebih kecil dari *threshold* yaitu 0.7. Dengan demikian kedua atribut tersebut dapat dianggap tidak serupa, sehingga tidak disarankan untuk menghapusnya.

Tabel 4. Contoh Bigram Proses pada Atribut Tidak Serupa

Basis data	Tabel	Atribut	Similaritas <i>Bigram</i>
Saduthama	property	keunggulan	ke,eu,un,ng,gg,gu,ul,la,an
SJP	property	deskripsi	de,es,sk,kr,ri,ip,ps,si
<i>Intersect</i>			0
<i>Similarity</i>			$= \frac{2 \times 0}{9 + 9}$
			$= 0$

Tabel 5 menunjukkan hasil *schema matching* antara basis data Sadhutama dan SJP. Proses ini dilakukan untuk membandingkan kesesuaian antar atribut dalam kedua basis data tersebut. Hasil pencocokan skema ini memberikan informasi mengenai tingkat kemiripan dan perbedaan antar elemen data. Jika ada atribut yang dianggap mirip, mereka bisa digabungkan atau diperlakukan sebagai duplikat. Sebaliknya, atribut yang sangat berbeda akan tetap dipisahkan untuk menjaga konsistensi dan keakuratan data.

Tabel 5. Hasil *Schema Matching*

Basis data				Bigram Similarity	Hasil Matching		
Saduthama		SJP					
Atribut	Tabel	Property	Atribut	Tabel	Property		
id			id			1	TP
id	fasilitas		id	fasilitas		1	TP
nama	property		name	property		0.9	TP
developer			developer			1	TP
tipe			type			0.8	TP
harga			harga			1.0	TP
luas tanah			luas tanah			1.0	TP
spesifikasi			spesifikasi			1.0	TP
lokasi			lokasi			1.0	TP
jenis property			jenis property			1.0	TP
kamar mandi			kamar mandi			1.0	TP
kamar tidur			kamar tidur			1.0	TP
ruangan keluarga			ruangan keluarga			1.0	TP
keunggulan			deskripsi			0	TN
alamat			jumlah lantai			0	TN

Hasil schema matching, yang diwakili oleh jumlah True Positives (TP), False Positives (FP), False Negatives (FN), dan True Negatives (TN), digunakan untuk mengevaluasi kinerja model schema matching. Nilai-nilai ini penting untuk mengukur akurasi dan efektivitas proses pencocokan. Tiga parameter utama digunakan untuk menilai kinerja model: precision (P), recall (R), dan F-Measure (F1 Score). Precision mengindikasikan proporsi atribut yang cocok dengan benar dari semua atribut yang dianggap cocok. Recall mengukur proporsi atribut yang cocok dengan benar dari total kecocokan yang relevan (lihat Persamaan (1)).

$$\begin{aligned}
 P &= \frac{TP}{TP+FP} \\
 P &= \frac{13}{13+0} \\
 P &= 1 \\
 R &= \frac{TP}{TP+FN} \\
 R &= \frac{13}{13+0} \\
 R &= 1 \\
 F \text{ Measure} &= \frac{2 \times P \times R}{P+R} \\
 F \text{ Measure} &= \frac{2 \times 1.0 \times 1.0}{1.0+1.0}
 \end{aligned} \tag{1}$$

Setelah proses *linguistic* selesai, langkah selanjutnya adalah *shema matching* dengan pendekatan *constraint*. Pada tahap ini, nilai 0 atau 1 diberikan berdasarkan kecocokan *constraint* antara sumber dan tujuan. Jika *constraint* pada sumber dan tujuan memiliki nilai yang sama, maka diberikan nilai 1. Sebaliknya, jika nilai *constraint* tersebut berbeda, maka diberikan nilai 0. Proses ini bertujuan untuk memastikan bahwa kesesuaian antara skema tidak hanya berdasarkan struktur, tetapi juga pada aturan atau batasan yang berlaku. Hasil *schema matching constraint* bisa dilihat pada Tabel 6.

Tabel 6. Contoh *Schema Matching* dengan Pendekatan *Constraint*

No	DB Utama (Saduthama)		Similarity		DB Target (SJP)			
	Atribut	Kriteria	Sama	Nilai	Kriteria	Atribut		
1	Nama property	Tipe	Int	Y	0.1	Int	Tipe	Name Property
		Lebar	20	Y	0.1	5	Lebar	
		Null	N	Y	0.1	N	Null	
		Unique	Y	Y	0.1	Y	Unique	
		Similarity Value				0.4		

Pada Tabel 7 nilai kemiripan yang ditunjukkan pada pasangan atribut adalah 0.4. Nilai kemiripan ini lebih besar dari *threshold* yang telah ditentukan, yaitu 0.3, yang menunjukkan adanya kesesuaian antara kedua atribut

tersebut. Dengan demikian, pasangan atribut ini dianggap relevan dan dapat diterima dalam proses *schema matching*. Oleh karena itu, pasangan atribut tersebut dikategorikan sebagai *true* positif. Hal ini menandakan bahwa pencocokan antar atribut tersebut valid dan sesuai dengan kriteria yang telah ditetapkan.

Tabel 7. Hasil *schema matching* dengan pendekatan *constraint*

Database Name		Bigram Similarity	Matching Results
Saduthama	SJP		
Atribut	Table Property	Atribut	Table Property
id	id	0.4	TP
id fasilitas	id fasilitas	0.4	TP
nama property	nama property	0.4	TP
developer	developer	0.4	TP
tipe	type	0.4	TP
harga	harga	0.4	TP
luas tanah	luas tanah	0.4	TP
spesifikasi	spesifikasi	0.4	TP
lokasi	lokasi	0.4	TP
jenis property	jenis property	0.4	TP
kamar mandi	kamar mandi	0.4	TP
kamar tidur	kamar tidur	0.4	TP
ruangan keluarga	ruangan keluarga	0.4	TP

Hasil dari pengujian pada Tabel 8 akan menjadi dasar untuk memperbaiki model pencocokan skema dan strategi integrasi basis data secara keseluruhan. Dengan memasukkan temuan-temuan ini ke dalam proses, kita dapat memastikan bahwa data dipetakan dengan akurat, mengurangi risiko kesalahan atau inkonsistensi. Proses perbaikan ini tidak hanya meningkatkan kualitas basis data yang terintegrasi, tetapi juga berkontribusi pada sistem manajemen data yang lebih kuat dan efisien, memastikan bahwa semua atribut ter-*align* dengan benar dan sistem yang terintegrasi berfungsi seperti yang diinginkan.

Tabel 8. Hasil presisi, *recall*, *F-measure* dan *accuracy* tiap *schema matching*

	Schema Matching 1		Schema Matching 2		Schema Matching 3	
	Linguistic	Constraint	Linguistic	Constraint	Linguistic	Constraint
Threshold	0.7	0.3	0.7	0.3	0.7	0.3
Presisi	84.1%	84.1%	94.7%	94.7%	94.7%	94.7%
Recall	84.1%	84.1%	94.7%	94.7%	94.7%	94.7%
F Measure	84.1%	84.1%	94.7%	94.7%	94.7%	94.7%
Accuracy	72%	72%	90%	90%	90%	90%

Tabel 9 menunjukkan hasil *schema matching* yang dilakukan antara empat basis data. Hasil ini mencakup perbandingan antara tabel-tabel dari masing-masing basis data. Setiap pencocokan skema dievaluasi menggunakan metrik seperti *precision*, *recall*, dan *F-measure*. Data yang ditampilkan pada tabel ini memberikan gambaran tentang sejauh mana atribut-atribut dalam basis data tersebut cocok satu sama lain. Hasil ini menjadi acuan untuk mengoptimalkan proses integrasi dan penggabungan data dari keempat basis data tersebut.

Tabel 9. Hasil *schema matching* atribut dari 4 basis data

Database	Table	Atribut
Saduthama	property	id, id fasilitas, nama property, developer, tipe harga, luas tanah, spesifikasi, lokasi, jenis property, kamar mandi, kamar tidur, ruangan keluarga, keunggulan, alamat
SJP	property	id, id fasilitas, name property, developer, type harga, luas tanah, deskripsi, lokasi, jenis, property, kamar mandi, kamar tidur, ruangan keluarga, spesifikasi, jumlah lantai
Combined Table	property	id, id fasilitas, name property, developer, type harga, luas tanah, deskripsi, lokasi, jenis property, kamar mandi, kamar tidur, ruangan keluarga, spesifikasi, jumlah lantai

3.3. Desain

Pada tahap ini, basis data didesain ulang berdasarkan hasil dari proses pencocokan skema. Proses pencocokan skema menghasilkan 8 tabel yang terdiri dari pengguna, fasilitas, dan properti. Tabel-tabel ini

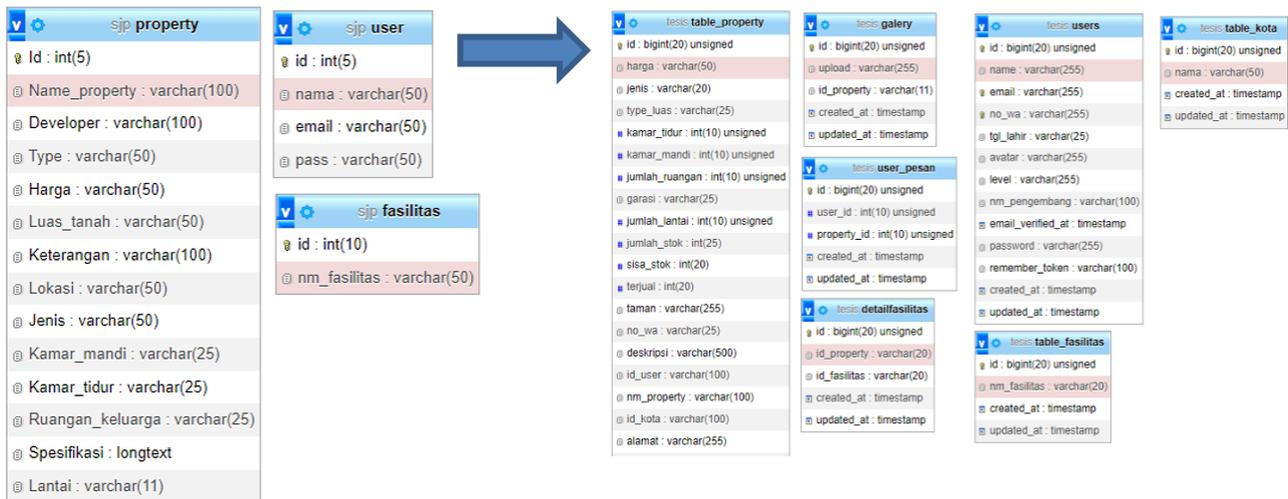
kemudian dimodelkan menjadi Diagram Entitas-Relasi (ERD). ERD selanjutnya dianalisis kembali berdasarkan kebutuhan pengguna untuk menambahkan entitas yang hilang. Pemodelan konseptual dibagi menjadi beberapa subsistem manajemen untuk memudahkan fokus pada setiap basis data: Saduthama, SJP, Salva, dan Garden View. Model konseptual untuk basis data ini melibatkan transformasi model basis data lama (hasil dari pencocokan skema) menjadi model baru. Transformasi ini mencakup penghapusan atribut redundan dan penambahan tabel baru. Sementara itu, model untuk Tabel Galeri, Detail Fasilitas, dan Lokasi adalah model baru yang sepenuhnya berdasarkan analisis kebutuhan. Dari model konseptual semua subsistem manajemen, perbandingan jumlah tabel dan atribut antara basis data lama dan baru ditampilkan pada Tabel 10.

Tabel 10. Perbandingan basis data lama dan baru

	Basis Data Lama	Basis Data Baru
Jumlah Tabel	12	8
Jumlah Atribut	80	41

3.4. Implementasi

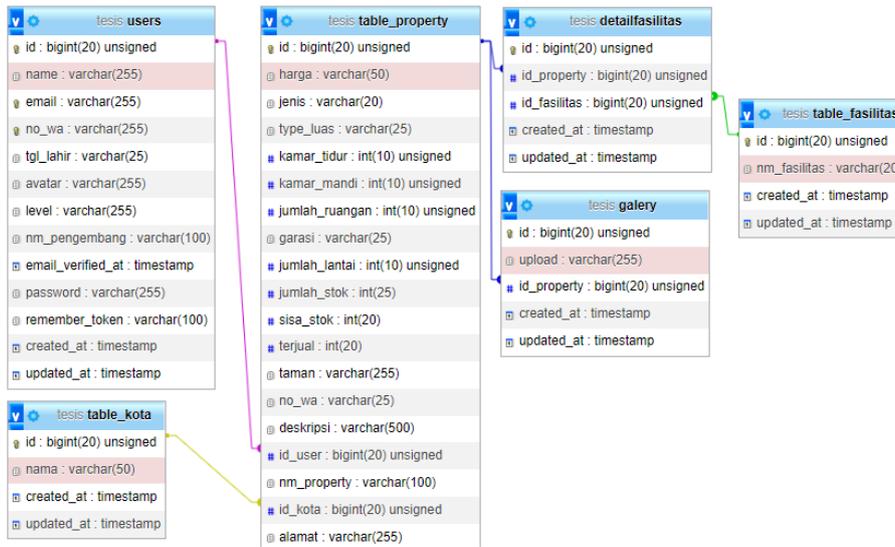
Pada tahap ini desain *Database* hasil pengembangan dapat diimplementasikan. Implementasi dilakukan dengan menggunakan *Database Management System (DBMS) My Structured Query Language (MySQL)*. Pemodelan logis memastikan bahwa setiap hubungan antar entitas sudah berbentuk *one-to-one* atau *one-to-many*, sehingga dapat diimplementasikan ke dalam basis data relasional. Selain itu, dilakukan validasi untuk memastikan bahwa bentuk normal hingga 3NF telah terpenuhi. Dari pemodelan konseptual yang telah dibuat, teridentifikasi bahwa setiap hubungan sudah memenuhi syarat relasional dan telah mencapai bentuk normal hingga 3NF. Tabel yang mencapai 3NF berarti tidak mengandung atribut nonkunci yang bergantung pada atribut nonkunci lainnya, sehingga semua atribut sepenuhnya bergantung pada primary key. Pemodelan logis juga memastikan bahwa setiap tabel yang terhubung telah dihubungkan dengan *foreign key* yang merujuk pada *primary key* dari tabel lain. Pemodelan fisik kemudian menerjemahkan pemodelan logis ini menjadi tabel yang siap diimplementasikan secara fisik, disesuaikan dengan DBMS yang dipilih (MySQL). Tahap ini mencakup pendefinisian teknis seperti nama tabel, nama kolom, tipe data, panjang data maksimal, nilai *default*, opsi *nullable*, dan atribut tambahan lainnya. Salah satu contoh hasil normalisasi dapat dilihat pada Gambar 3 Tahap ini mencakup pendefinisian teknis seperti nama tabel, nama kolom, tipe data, panjang data maksimal, nilai *default*, opsi *nullable*, dan atribut tambahan lainnya.



Gambar 3. Proses normalisasi

Gambar 4 menunjukkan sebelum normalisasi dan hasil setelah dilakukannya normalisasi. Dari gambar tersebut dapat dilihat bahwa setelah dilakukannya normalisasi dari tiga tabel menjadi enam tabel yaitu tabel

property, *user*, *user* pesan, fasilitas, detail fasilitas, kota dan galeri. Desain setelah normalisasi tersebut kemudian diubah dalam bentuk *physical*, karena desain yang ditunjukkan pada Gambar 3 adalah tahap *conceptual*. Bentuk *physical* desain dapat dilihat pada Gambar 4.



Gambar 4. Bentuk *physical*

Tabel 11 menunjukkan adanya perubahan jumlah sebelum dan setelah normalisasi pada tabel mahasiswa yang mana sebelumnya berjumlah 1 tabel berubah menjadi 4 tabel dan atributnya juga dari 16 menjadi 30. Secara keseluruhan hasil dari proses re-desain yang dilakukan dengan pengeleminasian dan normalisasi pada basis data Saduthama, SJP, Salva dan Garden View menyebabkan terjadinya perubahan jumlah tabel dan atribut yang dapat dilihat pada Tabel 11.

Tabel 11. Sebelum dan setelah normalisasi

	Sebelum Normalisasi	Setelah Normalisasi
Tabel	1	4
Atribut	16	30

Pada Tabel 12, dapat dilihat bagaimana perubahan jumlah tabel dan atribut memberikan dampak positif pada pengorganisasian data dan pengurangan redundansi. Data yang ditampilkan menunjukkan bahwa proses schema matching dan normalisasi berhasil meningkatkan struktur dan efisiensi basis data. Dan menunjukkan adanya perubahan jumlah tabel dan atribut untuk masing-masing basis data.

Tabel 12. Hasil redesain

Basis Data	Sebelum re-desain		Setelah re-desain	
	Tabel	Atribut	Tabel	Atribut
Saduthama	3	21	8	41
SJP	3	21		
Salva	3	19		
Garden View	3	19		

4. KESIMPULAN

Penelitian ini berhasil mengimplementasikan schema matching untuk mengintegrasikan basis data properti dari empat sumber berbeda: Saduthama, Salva, SJP, dan Garden View. Proses integrasi mencakup preprocessing, generalisasi, dan matching, dengan tujuan menyalarkkan atribut-atribut pada masing-masing basis data. Metode linguistic dengan teknik Bigram menunjukkan efektivitas dalam mengidentifikasi atribut serupa berdasarkan tingkat kemiripan string, dengan nilai threshold 0,7. Hasil analisis menunjukkan bahwa atribut dengan nilai

kemiripan $\geq 0,7$ dapat dianggap sama, sehingga atribut duplikat dapat dihapus untuk mengurangi redundansi. Pada tahap selanjutnya, constraint matching mengevaluasi kesesuaian atribut berdasarkan kriteria tambahan seperti tipe data, panjang atribut, nilai null, dan unik. Nilai kesesuaian dihitung menggunakan bobot 0,1 untuk setiap kriteria yang cocok, memberikan gambaran lebih rinci mengenai kesamaan antar atribut.

Dibandingkan dengan penelitian sebelumnya oleh [11] yang menghasilkan nilai presisi 100%, recall 60%, dan F-measure 75%, penelitian ini memberikan peningkatan dengan nilai rata-rata presisi (P), recall (R), dan F-measure sebesar 90%, serta rata-rata akurasi sebesar 84%. Hal ini menunjukkan bahwa metode schema matching yang diusulkan lebih efektif dan akurat dalam mencocokkan atribut antar basis data.

Namun, penelitian ini memiliki keterbatasan pada metode Bigram, yang kurang optimal saat diterapkan pada atribut dengan jumlah karakter sedikit. Keterbatasan ini menyebabkan akurasi pencocokan atribut menjadi kurang maksimal pada string pendek. Sebagai saran untuk penelitian selanjutnya, disarankan untuk mempertimbangkan penggunaan metode pencocokan string alternatif, seperti Levenshtein Distance atau Jaccard Similarity, untuk meningkatkan akurasi pada atribut dengan string pendek. Selain itu, pengembangan metode hybrid dengan menggabungkan pendekatan linguistik dan constraint secara dinamis dapat memberikan hasil yang lebih adaptif terhadap berbagai jenis data.

UCAPAN TERIMA KASIH

Kami ingin mengucapkan terima kasih kepada Bapak Bambang Krismono Triwijoyo sebagai pembimbing pertama dan Bapak Galih Hendro Martono sebagai pembimbing kedua atas bimbingan dan dukungan mereka selama penelitian ini.

DAFTAR PUSTAKA

- [1] J. Okto dan S. Hendraputra, "Perancangan Sistem informasi Pemasaran Rumah pada PT.Nakama Berbasis Web dengan Menggunakan Metode Waterfall," *REMIK: Riset dan E-Jurnal Manajemen Informatika Komputer*, vol. 6, no. 2, pp. 304–317, Jul. 2022. DOI: [10.33395/remik.v6i2.11547](https://doi.org/10.33395/remik.v6i2.11547).
- [2] I. H. Primananda, N. Santoso, dan A. Rachmadi, "Pengembangan Sistem Marketplace Tanah dan Property Berbasis Web dengan Pendekatan Rapid Application Development (RAD)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 4, no. 1, pp. 200–206, Feb. 2020.
- [3] D. N. Kholifah *et al.*, "Perancangan Program Absensi Karyawan Berbasis Web Menggunakan Metode Waterfall Pada PT Kedai Sayur Indonesia," *Indonesian Journal on Software Engineering (IJSE)*, vol. 8, no. 1, pp. 115–124, Jun. 2022. DOI: [10.31294/ijse.v8i1.13025](https://doi.org/10.31294/ijse.v8i1.13025).
- [4] O. Prasetyo dan W. G. Aedi, "Implementasi Teknologi Augmented Reality Untuk Bisnis Properti Rumah Menggunakan Metode Markerless Berbasis Android (Studi Kasus CV. Tomsland)," *Jurnal Informatika MULTI*, vol. 1, no. 4, pp. 275–284, 2023.
- [5] A. Nurzukhrufa, W. D. Satria, dan M. K. Dewi, "Strategi Pemasaran Pengembang Perumahan Saat Pandemi Covid-19 di Kota Bandar Lampung," *Jurnal Maksipreneur: Manajemen, Koperasi, dan Entrepreneurship*, vol. 12, no. 1, p. 45, Jul. 2022. DOI: [10.30588/jmp.v12i1.882](https://doi.org/10.30588/jmp.v12i1.882).
- [6] S. Raharjo, E. Utami, dan O. M. A. Alsyabani, "Metode Schema Matching berbasis Linguistic dan Constraint untuk Integrasi Database di Sekolah," *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 8, no. 2, p. 311, Aug. 2022. DOI: [10.26418/jp.v8i2.55852](https://doi.org/10.26418/jp.v8i2.55852).
- [7] M. Shrestha *et al.*, "Schema Matching and Data Integration with Consistent Naming on Protein Crystallization Screens," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 17, no. 6, pp. 2074–2085, Nov. 2020. DOI: [10.1109/TCBB.2019.2913368](https://doi.org/10.1109/TCBB.2019.2913368).

- [8] K. Christianto *et al.*, “Pengembangan Sistem Aplikasi Manajemen Properti Bagian Penyewaan Unit (Studi Kasus: PT. Satrand Group Indonesia),” *Infotech: Journal of Technology Information*, vol. 9, no. 1, pp. 37–42, Jun. 2023. DOI: [10.37365/jti.v9i1.156](https://doi.org/10.37365/jti.v9i1.156).
- [9] H. Gunawan dan A. Maslan, “Aplikasi Promosi Properti Berbasis Web Menggunakan Metode Waterfall,” *Computer and Science Industrial Engineering (COMASIE)*, vol. 6, no. 2, pp. 80–87, Jan. 2022.
- [10] D. A. Herman, “Perancangan dan Pengembangan Aplikasi Jual Beli Properti Berbasis Web dengan Metode SDLC (Studi Kasus: PT Limitra Indonesia Mega),” *Journal of Information System and Technology*, vol. 3, no. 2, p. 35, Jul. 2022. DOI: [10.37253/joint.v3i2.6755](https://doi.org/10.37253/joint.v3i2.6755).
- [11] A. E. Permanasari *et al.*, “Pengembangan Basis Data Sistem Informasi Manajemen Rumah Sakit Berbasis Linguistic-based Schema Matching,” *Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI)*, vol. 8, no. 2, p. 101, May 2019. DOI: [10.22146/jnteti.v8i2.498](https://doi.org/10.22146/jnteti.v8i2.498).
- [12] R. Hammad *et al.*, “Optimization of data integration using schema matching of linguistic-based and constraint-based in the university database,” *Matrix : Jurnal Manajemen Teknologi dan Informatika*, vol. 11, no. 3, pp. 119–129, Nov. 2021. DOI: [10.31940/matrix.v3i11.119-129](https://doi.org/10.31940/matrix.v3i11.119-129).
- [13] A. Gal, H. Roitman, dan R. Shraga, “Learning to Rerank Schema Matches,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 8, pp. 3104–3116, Aug. 2021. DOI: [10.1109/TKDE.2019.2962124](https://doi.org/10.1109/TKDE.2019.2962124).
- [14] C. Ma *et al.*, “Knowledge Enriched Schema Matching Framework for Heterogeneous Data Integration,” in *2022 IEEE 2nd Conference on Information Technology and Data Science (CITDS)*, Debrecen, Hungary: IEEE, May 2022, pp. 183–188. DOI: [10.1109/CITDS54976.2022.9914350](https://doi.org/10.1109/CITDS54976.2022.9914350).
- [15] X. Du *et al.*, “RESIN-11: Schema-guided Event Prediction for 11 Newsworthy Scenarios,” in *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: System Demonstrations*, Hybrid: Seattle, Washington + Online: Association for Computational Linguistics, 2022, pp. 54–63. DOI: [10.18653/v1/2022.naacl-demo.7](https://doi.org/10.18653/v1/2022.naacl-demo.7).
- [16] M. Tamer Özsü dan P. Valduriez, “Correction to: Principles of Distributed Database Systems,” in *Principles of Distributed Database Systems*. Cham: Springer International Publishing, 2020, pp. C1–C2. DOI: [10.1007/978-3-030-26253-2_13](https://doi.org/10.1007/978-3-030-26253-2_13).
- [17] I. R. Mukhlis dan R. Santoso, “Perancangan Basis Data Perpustakaan Universitas Menggunakan MySQL dengan Physical Data Model dan Entity Relationship Diagram,” *Journal of Technology and Informatics (JoTI)*, vol. 4, no. 2, pp. 81–87, Apr. 2023. DOI: [10.37802/joti.v4i2.330](https://doi.org/10.37802/joti.v4i2.330).

[Halaman ini sengaja dikosongkan.]