Waii Home About

Unlocking the Power of Text-to-SQL: How Metadata Makes the Difference

Gunther Hagleitner · Follow

Published in Waii · 6 min read · Mar 5, 2025

In the evolving landscape of data analytics, the ability to simply ask questions of your data and receive accurate SQL queries in return is no longer science fiction — it's a competitive advantage. Waii stands at the forefront of this revolution, offering a text-to-SQL solution that truly understands your data's context. Let's explore how Waii's multi-faceted approach to metadata transforms conversational analytics from frustrating to fantastic.

The Secret Ingredient: Metadata Magic

What separates good text-to-SQL from great text-to-SQL? In a word: metadata. Without proper context, even the most sophisticated AI will struggle to translate natural language into accurate queries. Waii's approach tackles this challenge head-on with a comprehensive strategy that can quickly achieve 95%+ accuracy.

Automatic Semantic Context Generation: The Foundation

Imagine setting up a new analytics tool and discovering it already understands your database intimately. That's the reality with Waii's automatic semantic context generation. By examining your database structure, Waii can:

- Predict joins between tables
- Categorize columns as measures, metrics, dimensions, or hierarchies
- Analyze data patterns for similarity search
- Identify substructures and schema relationships
- Pinpoint high-value content
- Infer data lineage
- Predict high-value metrics and KPIs
- Formulate business context for each object

This isn't just AI — it's AI combined with sophisticated compiler technology that understands database structures at a fundamental level. For well-modeled databases with clear documentation, this first step alone can deliver surprising accuracy.

Leveraging Your Existing Documentation

Why reinvent the wheel? Waii automatically ingests the documentation and metadata you've already created:

- Technical documentation (e.g.: text, excel, websites, pdf, confluence)
- Semantic layers (e.g.: LookerML, dbt)

- Business glossaries (e.g.: Collibra, Alation)
- Data catalogs (e.g.: OpenMetadata.org, DataHub)

Waii doesn't just store this information — it intelligently links each concept to the right nodes in its knowledge graph, creating a rich understanding of your specific domain.

The agents used to build your SQL query dynamically retrieve these linked concepts and will prioritize them over automatically generated context — when you spend the time to model a concept, Waii makes sure to follow the instructions.

Need to refine something quickly? Simply add incidental documentation to clarify concepts, and watch as Waii's understanding evolves.

360° Evaluation: Testing for Excellence

Great systems aren't built in a vacuum — they're tested rigorously. Waii has partnered with Microsoft to implement Archerfish (https://github.com/archerfish-bench), an evaluation framework that measures performance based on user-intent metrics rather than just syntactic accuracy.

To ensure comprehensive coverage, Waii generates question/answer pairs that reflect real-world usage patterns:

- Automatically created from your environment
- Informed by your query history
- Based on existing dashboards and reports

After review, these datasets train and tune the system, eliminating remaining shortcomings and ensuring balanced performance across all use cases.

Continuous Learning: Getting Better Every Day

The magic doesn't stop at deployment. Waii's live feedback system captures user interactions — correct responses, incorrect ones, and explicit feedback — to continuously refine its understanding.

Here's how it works: When feedback is received, Waii evaluates it against its knowledge graph, identifying precisely what needs to be adjusted. Then, when a similar question arises, Waii applies these learnings as a delta to ensure the feedback is respected in future responses.

Importantly, you maintain full control over how this feedback propagates through your organization. Administrators can approve which feedback gets distributed to which users or teams, ensuring that improvements benefit the right people while maintaining consistent experiences across different groups when needed.

This creates a virtuous cycle where every interaction makes the system smarter — but in a controlled, intentional way that respects your organization's structure and needs.

Organization That Makes Sense

How does Waii manage all this metadata? Through a sophisticated labeling system that tracks:

- Information origin

- Date of changes
- Modification history
- Multi-tenant capabilities

Different teams can maintain different understandings of concepts, and administrators can implement approval processes to control access and changes to semantic context.

Version Control Ready: Store Your Metadata as Code

Waii takes a modern approach to metadata management with its YAML-based language that allows you to store all semantic information as code in GitHub. This brings powerful advantages:

- Track changes with familiar git workflows
- Review metadata modifications through pull requests
- Roll back to previous versions when needed
- Collaborate across teams with branching strategies
- Automate deployment through CI/CD pipelines

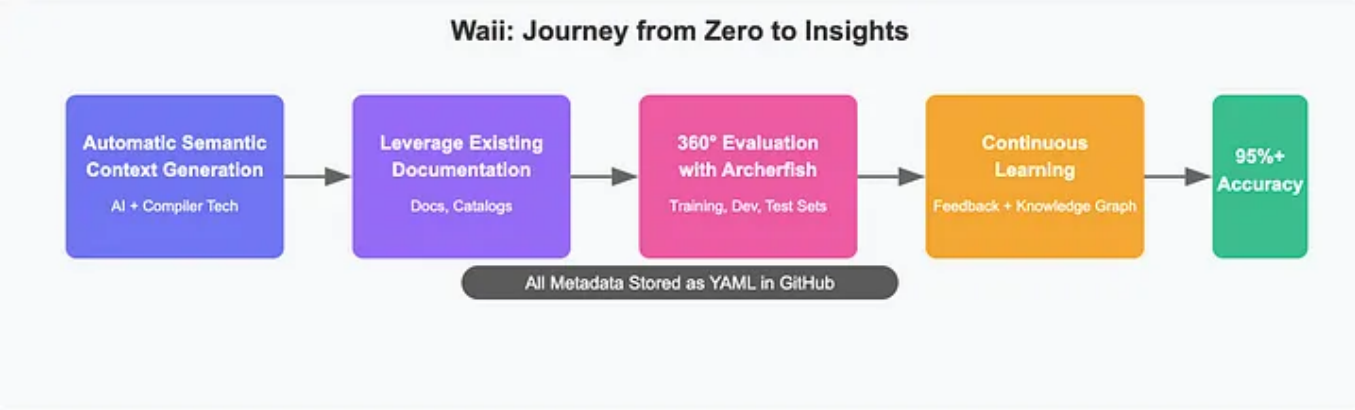
Here's what this looks like in practice:

```
tables:
  - name: orders
    description: "Tracks all customer orders with payment and fulfillment status
    pk: order_id
    semantic_context:
      - "Queries against this table need to filter by load_id for proper SCD"
    joins:
```

```
    - description: customers
      type: N:1
      fk: customer_id
      constraints: "orders.customer_id = customers.id"
  columns:
    - name: product_category
      type: categorical
      vector_search: product_embeddings.vectors
  measures:
    - name: revenue
      expr: SUM(amount)
      description: "Total order value in USD"
```

This code-first approach to metadata means your semantic layer becomes a first-class asset in your development ecosystem — versioned, reviewed, and deployed with the same rigor as your application code.

From Zero to Insights in Record Time



The true testament to Waii's approach isn't just its technical sophistication — it's the speed of implementation. Typically, customers go from nothing to a working system in less than 24 hours.

Although varying with business requirements and how much of a custom UX is desired, it's common to start onboarding users just a couple of weeks after

first setup.

This rapid time-to-value means you're not waiting months for benefits. Your team can start asking questions in natural language and receiving accurate SQL fast, creating immediate impact while the system continues to refine its understanding through usage.

The Result: A System That Speaks Your Language

When these approaches combine, something magical happens: your analytics tool suddenly speaks your language. It understands your business definitions, recognizes your domain terminology, and translates natural language into precisely the SQL you would have written yourself.

See It In Action

LookML Metric Definition:

```
measure: customer_retention_rate {
  type: number
  sql: ${retained_customers} / NULLIF(${previous_customers}, 0) ;;
  value_format_name: percent_2
  description: "Percentage of customers who made a purchase in previous period a
}
```

User Question: "What was our customer retention for Q1 by region?"

Waii Generated SQL:

```
SELECT
  region,
  COUNT(CASE WHEN purchased_prev_quarter = 1 AND purchased_this_quarter = 1 THEN
  NULLIF(COUNT(CASE WHEN purchased_prev_quarter = 1 THEN customer_id END), 0) AS
FROM customer_activity
WHERE quarter = '2025-Q1'
GROUP BY region
ORDER BY customer_retention_rate DESC
```

User Feedback: “This is good but should include the rate and the actual counts”

Waii Learning:
✓ Associate “customer retention” with the specific LookML measure
✓ Remember to include component counts when this user asks for rates

Next Time Query:

```
SELECT
  region,
  COUNT(CASE WHEN purchased_prev_quarter = 1 THEN customer_id END) AS previous_c
  COUNT(CASE WHEN purchased_prev_quarter = 1 AND purchased_this_quarter = 1 THEN
  COUNT(CASE WHEN purchased_prev_quarter = 1 AND purchased_this_quarter = 1 THEN
  NULLIF(COUNT(CASE WHEN purchased_prev_quarter = 1 THEN customer_id END), 0) AS
FROM customer_activity
WHERE quarter = '2025-Q1'
GROUP BY region
ORDER BY customer_retention_rate DESC
```

In a world where data drives decisions, the ability to simply ask questions of your data — and get accurate answers immediately — isn’t just convenient.

It's transformative. And with Waii's metadata-driven approach to text-to-SQL, that transformation is within reach.

Ready to experience the difference metadata makes? Discover how Waii can bring conversational analytics to your organization today.

Generative Ai Tools | Text To Sql | Generative Bi | Metadata | Sql Agent

Published in Waii

97 Followers · Last published Mar 5, 2025

Life at the intersection of AI and Data

Follow

Written by Gunther Hagleitner

118 Followers · 42 Following

Data nerd and pilot from the SF bay area.

Follow

No responses yet

S Java

What are your thoughts?

More from Gunther Hagleitner and Waii

In Waii by Gunther Hagleitner

Complex SQL Joins with LangGraph and Waii

In the rapidly evolving landscape of data analytics, the ability to interact with data...

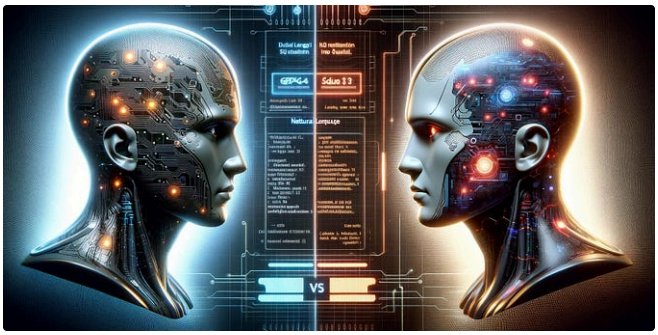
Oct 11, 2024 133 2

In Waii by Rachana Aluri

Text-to-SQL's Power Players: Comparing Claude 3.5 Sonnet,...

The ability to query databases using natural language continues to revolutionize data...

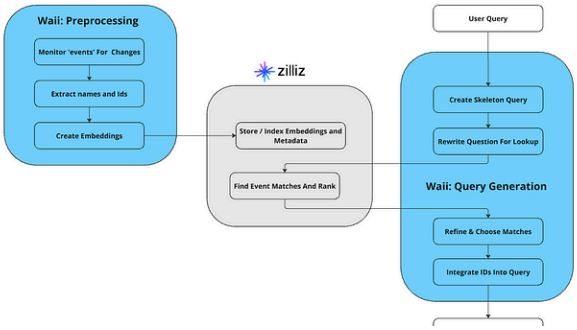
Sep 4, 2024 220



In Waii by Rajesh Balamohan

Translating Natural Language into SQL: GPT-4 vs Claude 3 Showdown

Anthropic recently released the Claude 3 model (Introducing the next generation of...



In Waii by Gunther Hagleitner

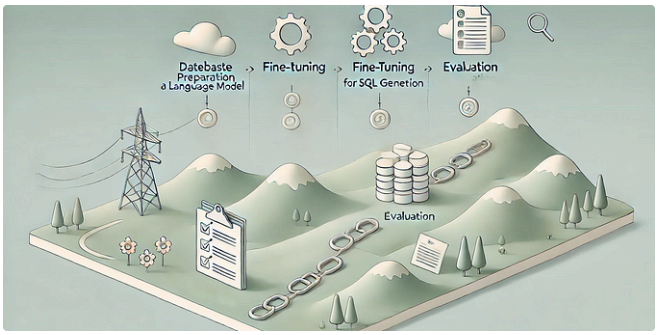
Tame High-Cardinality Categorical Data in Agentic SQL Generation...

By Jiang Chen and Gunther Hagleitner

Mar 15, 2024 70 1 Sep 16, 2024 159

See all from Gunther Hagleitner See all from Waii

Recommended from Medium



Yi Ai

Enhancing Text-to-SQL with a Fine-Tuned 7B LLM for Database...

A Step-by-Step Guide to Fine-Tuning Models for SQL Generation on Custom Database...

Sep 29, 2024 101 1

Justin Muller

The biggest problem with Text to SQL workloads, and how to fix it.

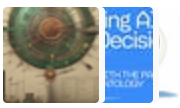
Designing for the human element of Text to SQL workloads.

Jan 6 19 1

Lists

Business

41 stories · 183 saves



data science and AI

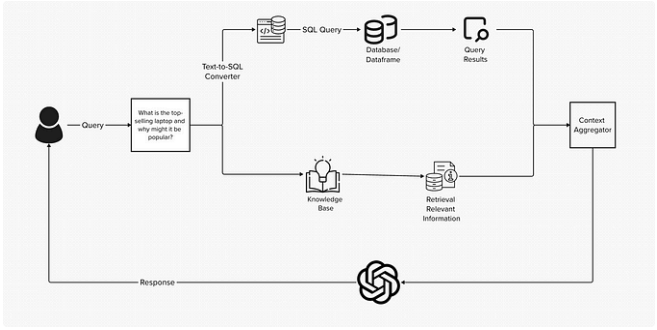
40 stories · 341 saves

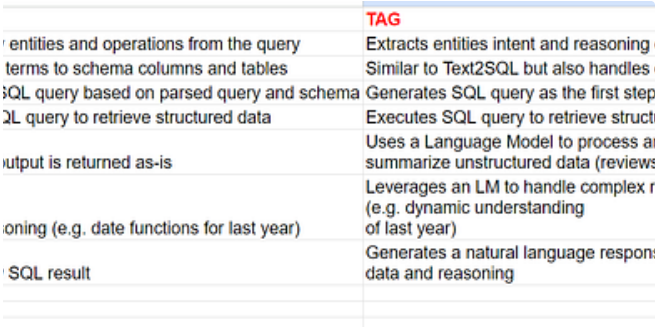
Natural Language Processing

1977 stories · 1620 saves

Medium's Huge List of Publications Accepting...

414 stories · 4679 saves





	TAG
entities and operations from the query	Extracts entities intent and reasoning
terms to schema columns and tables	Similar to Text2SQL but also handles
QL query based on parsed query and schema	Generates SQL query as the first step
L query to retrieve structured data	Executes SQL query to retrieve struct
utput is returned as-is	Uses a Language Model to process a summarize unstructured data (reviews
oning (e.g. date functions for last year)	Leverages an LM to handle complex r (e.g. dynamic understanding of last year)
SQL result	Generates a natural language respons data and reasoning

Aakriti Aggarwal

AI’nt That Easy #18: Text-to-SQL, RAG, and TAG Explained

Four approaches — Text-to-SQL, RAG, RAG over Tabular Data, and TAG — each offer...

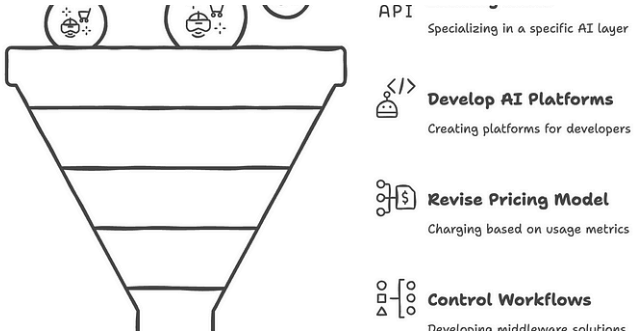
Oct 5, 2024 · 109 · 1

In Data Scientists from Future by Vishal Mysore

TAG vs Text2SQL

To understand how TAG (Table-Augmented Generation) and Text2SQL differ internally in...

Sep 18, 2024 · 73



Julio Pessan

Don’t Sell AI Agents, Sell AI Infrastructures Instead—The...

The AI Mirage—And the Fortune Few See Coming

6d ago 535 26

In Data Science Collective by Eric Broda

Agentic Mesh: Building Highly Reliable Agents

LLMs are getting overloaded. Specialized LLMs, with deterministic orchestration & an...

Mar 5 480 8

See more recommendations