

ASMR: Agentic Schema Generation for Ship Maintenance Report Writing

Sohrab Namazi Nia
New Jersey Institute of Technology
Newark, New Jersey, USA
sn773@njit.edu

Peter Ly
Rensselaer Polytechnic Institute
Troy, New York, USA
lyp2@rpi.edu

Jay Miller
Boston Fusion Corporation
Lexington, Massachusetts, USA
jay.miller@bostonfusion.com

Amogh Dalal
New Jersey Institute of Technology
Newark, New Jersey, USA
asd33@njit.edu

Marti Zentmaier
Boston Fusion Corporation
Lexington, Massachusetts, USA
marti.zentmaier@bostonfusion.com

Rishi Singh
Boston Fusion Corporation
Lexington, Massachusetts, USA
rishi.singh@bostonfusion.com

Ning Sa
Rensselaer Polytechnic Institute
Troy, New York, USA
san2@rpi.edu

Tomek Strzalkowski
Rensselaer Polytechnic Institute
Troy, New York, USA
tomek@rpi.edu

Senjuti Basu Roy
New Jersey Institute of Technology
Newark, New Jersey, USA
senjutib@njit.edu

ABSTRACT

In this paper, we study the *automatic schema generation problem*: given a collection of historical ship maintenance and operational reports across multiple form categories, automatically discover compact and informative schemas that capture the essential information requirements of each report type. To address this challenge, we propose **ASMR**, a modular agentic framework consisting of two specialized agents. A *Field Generation Agent* extracts semantic concepts from historical narratives and generates candidate schema fields through adaptive multi-granularity clustering, while a *Structural Optimizer Agent* employs reinforcement learning to identify compact, informative, and non-redundant schema representations. The resulting schemas can guide report authors toward producing more complete, consistent, and actionable reports. Preliminary results demonstrate the promise of the proposed approach and highlight several open research challenges at the intersection of data management, agentic AI, and human-centered AI.

VLDB Workshop Reference Format:

Sohrab Namazi Nia, Amogh Dalal, Ning Sa, Peter Ly, Marti Zentmaier, Tomek Strzalkowski, Jay Miller, Rishi Singh, and Senjuti Basu Roy. ASMR: Agentic Schema Generation for Ship Maintenance Report Writing. VLDB 2026 Workshop: DASHSys: Systems for Data-centric Agents with Human-in-the-loop.

1 INTRODUCTION

Many ship operational and maintenance reporting workflows rely on forms that contain substantial unstructured textual narratives. Free-form text provides flexibility to describe critical operational

information such as equipment conditions, hazards, mitigation actions, system failures, preservation activities, maintenance observations, and operational impacts. However, across different personnel with varying levels of expertise, this same flexibility can result in inconsistent structure that makes it difficult to extract actionable insights from reports at scale. In particular, required information that goes unrecorded by personnel can be impossible to recover, which is a problem made worse in large enterprises where report writers are far removed from downstream report consumers. Instead of burdening personnel with more training that scales poorly across large workforces, we imagine real-time writing assistance that provides topic-specific information requirements to personnel during the writing process. This would better optimize trade-offs between flexibility and structure provided by free-text inputs while supporting downstream workflows.

To support such a capability, we formulate the *schema generation problem*: given a collection of historical reports, automatically discover schemas consisting of compact sets of information fields that capture essential concepts and report information requirements; additionally, as the relevant concepts and information requirements for a report will change with its general topic, each schema should correspond to a specific report topic. A schema field represents a recurring semantic aspect of the reports, such as *Condition*, *Hazard*, *Mitigation*, *Location*, or *System Failure*. The goal is to discover schema fields that are sufficiently expressive to capture the diversity of historical reports while remaining concise, interpretable, and minimally redundant.

To address this challenge, we study the automatic schema generation problem and propose **ASMR**, a modular agentic framework that combines Large Language Model (LLM)-based semantic understanding with reinforcement learning-based optimization techniques. The framework consists of two specialized agents, as illustrated in Figure 1. The first, a *Field Generator Agent*, analyzes historical reports, extracts salient semantic concepts from unstructured narratives, adaptively clusters related concepts at multiple granularities, and generates candidate schema fields. The second,

This work is licensed under the Creative Commons BY-NC-ND 4.0 International License. Visit <https://creativecommons.org/licenses/by-nc-nd/4.0/> to view a copy of this license. For any use beyond those covered by this license, obtain permission by emailing info@vldb.org. Copyright is held by the owner/author(s). Publication rights licensed to the VLDB Endowment.
Proceedings of the VLDB Endowment. ISSN 2150-8097.

a *Structural Optimizer Agent*, models schema construction as a sequential decision-making problem and employs reinforcement learning-based optimization to iteratively retain, merge, or discard candidate fields in order to produce compact, informative, and non-redundant schemas. The key insight is that schema design can be formulated as a multi-objective optimization problem that balances coverage, informativeness, compactness, and redundancy.

ASMR enables AI-assisted reporting by generating structured schemas that improve report completeness and consistency while supporting downstream analytics such as predictive maintenance [17], operational readiness assessment, anomaly detection, and knowledge extraction. Since the framework learns directly from historical reports, it can also adapt to evolving reporting requirements with minimal manual redesign.

Although **ASMR** is motivated by and evaluated on ship operational and maintenance forms, we believe it has applicability to a broad range of industrial, manufacturing, healthcare, logistics, aviation, and infrastructure management systems that rely heavily on semi-structured and unstructured forms and face similar challenges. Development of **ASMR** is currently in progress, and our initial evaluations provide encouraging evidence. Compared to recent systems such as PALIMPZEST [8], ELEET [15], DocETL [14], and TabAgent [18], the proposed work addresses a complementary problem centered on automatic schema generation for human-AI collaborative operational reporting.

Contributions.

- We formulate the *automatic schema generation problem* for operational and maintenance reporting and propose **ASMR**, a modular agentic framework that combines LLM-based semantic concept extraction, adaptive multi-granularity clustering, and reinforcement learning-based optimization to automatically generate compact, informative, and non-redundant schemas from historical reports.
- We demonstrate how the generated schemas can enable structured human-AI collaboration by automatically producing field-specific guidance and questions that assist report authors in creating more complete, consistent, and actionable operational reports.
- We develop quantitative metrics for evaluating schema quality and demonstrate through preliminary experiments that **ASMR** produces compact, informative, and low-redundancy schemas that support effective AI-assisted operational reporting.

The rest of the paper is organized as follows: in Section 2, we present our data model and formalize the proposed problem. Section 3 presents the proposed modular agentic framework. Section 4 presents preliminary experimental evaluation and results. Related work is discussed in Section 5. We discuss challenges and open problems in Section 6 and conclude in Section 7.

2 DATA MODEL & PROBLEM DEFINITION

Historical Forms. Let $\mathcal{H} = \{f_1, f_2, \dots, f_n\}$ denote a collection of historical maintenance forms spanning multiple operational categories. Each form instance $f_i \in \mathcal{H}$ corresponds to a free-form

narrative associated with a form type t , such as *VOIDS AND COFFERDAMS*, *LIVING QUARTERS COMPARTMENTS*, *STORAGE COMPARTMENTS*, or *FUEL OIL TANKS*. These narratives contain heterogeneous operational observations, maintenance descriptions, hazards, mitigation procedures, and equipment-related information expressed in unstructured natural language.

Concepts. A concept represents a semantically meaningful atomic unit extracted from a historical form narrative. Formally, given a form instance f_i , a concept $c \in C$ corresponds to an identifiable semantic entity, operational descriptor, condition, or event appearing in the narrative. Examples of concepts extracted from the forms include *corrosion*, *water intrusion*, *improper sealing*, *water separator assemblies*, *degraded lagging*, etc. Concepts are generated by the Field Generator Agent using an LLM-driven semantic extraction mechanism and capture the fine-grained semantics present in historical reports.

Fields. A field represents a structured attribute generated by aggregating semantically related concepts across multiple historical forms. Formally, a field $\phi \in \Phi$ corresponds to a canonical abstraction over a subset of concepts $C_\phi \subseteq C$ that share operational or semantic similarity. For example, concepts such as *Leakage*, *Leak*, and *System Failure* may map to a failure-related field, while concepts such as *Location*, *Accessibility*, and *Bulkhead* may map to a compartment-related field. The resulting field space Φ defines the candidate field space generated by the Field Generator Agent and subsequently optimized into schemas for different form types.

Schema of a Form Type. Given a form type t , its schema $\mathcal{S}_t$ is defined as a structured collection of fields that collectively characterize the semantic and operational information expected to appear in forms of type t . Formally,

$$\mathcal{S}_t = \{\phi_1, \phi_2, \dots, \phi_m\},$$

where each $\phi_i \in \Phi$ represents a canonical field generated by aggregating semantically related concepts extracted from historical narratives belonging to form type t . The schema captures the recurring semantic structure underlying a category of operational reports.

For example, the schema corresponding to the form type *FUEL OIL TANKS* may contain fields such as

$$\mathcal{S}_{\text{FuelOilTanks}} = \left\{ \begin{array}{l} \text{Condition,} \\ \text{Location,} \\ \text{FailureDescription,} \\ \text{ElectricalSystems,} \\ \text{Mitigation} \end{array} \right\}$$

where the field *FailureDescription* may aggregate concepts such as *Leakage*, *Leak*, and *System Failure*, while *ElectricalSystems* may aggregate concepts such as *Electrical* and *Electronics*. Similarly, the schema for *VOIDS AND COFFERDAMS* may contain fields such as *Location*, *Hazard*, *Condition*, *Accessibility*, and *Preservation*.

Problem Definition. Given a collection of historical form narratives

$$\mathcal{H} = \{f_1, f_2, \dots, f_n\},$$

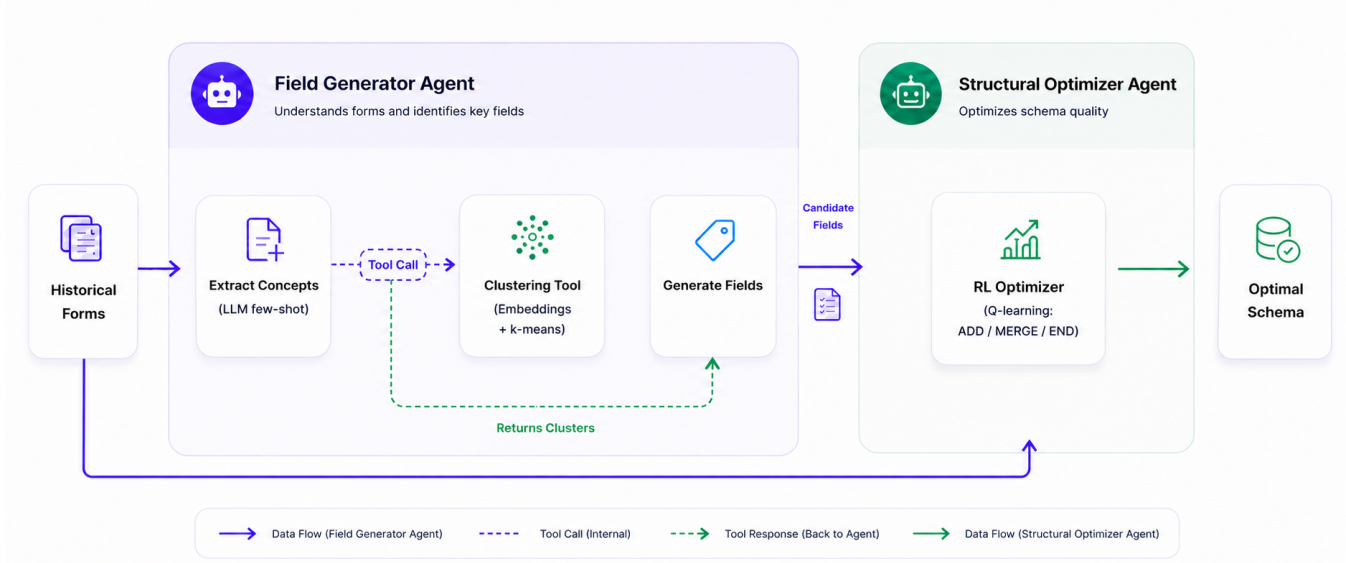


Figure 1: Overview of the proposed ASMR framework consisting of a Field Generator Agent for semantic concept extraction and candidate field generation, and a Structural Optimizer Agent for RL-based schema optimization.

where each form instance $f_i \in \mathcal{H}$ belongs to one of m different form types

$$\mathcal{T} = \{t_1, t_2, \dots, t_m\},$$

and each instance consists of unstructured free-form operational text, the objective is to automatically discover an optimal structured schema

$$\mathcal{S}_{t_j} = \{\phi_1, \phi_2, \dots, \phi_k\}$$

for every form type $t_j \in \mathcal{T}$ such that the schema captures the recurring semantic structure underlying the historical forms belonging to that form type.

Our proposed agentic framework addresses this problem through two specialized agents:

- (i) a *Field Generator Agent* that extracts a set of semantic concepts

$$\mathcal{C} = \{c_1, c_2, \dots, c_r\}$$

from historical narratives, groups semantically related concepts to consolidate overlapping information across multiple forms and form types, and generates an overcomplete set of candidate fields

$$\Phi = \{\phi_1, \phi_2, \dots, \phi_p\},$$

and

- (ii) a *Structural Optimizer Agent* that optimizes the candidate field space to construct the final schema $\mathcal{S}_{t_j}$ for each form type t_j .

3 ASMR

ASMR is a modular agentic framework for automatic schema generation from historical ship maintenance and operational forms. It mainly consists of two specialized agents: a Field Generator Agent that extracts semantic concepts and generates an overcomplete candidate field space, and a Structural Optimizer Agent that applies

reinforcement learning-based optimization to construct compact, informative, and non-redundant schemas for different form types.

Figure 1 presents an overview of the overall ASMR framework. Section 3.1 describes the Field Generator Agent, while Section 3.2 presents the Structural Optimizer Agent. To quantitatively assess the quality of the generated schemas by ASMR, we further introduce our schema quality evaluation methodology, described in Section 3.3.

3.1 Field Generator Agent

The Field Generator Agent is an LLM-based agent that discovers and abstracts semantic information requirements from historical maintenance narratives. As illustrated in Figure 2, the agent extracts semantic concepts from unstructured reports, leverages a clustering tool to organize related concepts into coherent groups, and generates an overcomplete candidate field space. The resulting candidate fields are then provided to the Structural Optimizer Agent for schema optimization.

3.1.1 Semantic Concept Extraction. Historical forms are first grouped according to their corresponding form types using ESWBS codes (Engineering Systems Breakdown Structure identifiers used to categorize operational and maintenance components). For each form type $t \in \mathcal{T}$, representative subsets of historical forms are sampled from $\mathcal{H}$ and provided to the LLM as the core semantic engine of the Field Generator Agent.

The concept extraction process relies on instruction-guided few-shot prompt engineering, where sampled historical forms and task-specific parameters, such as the number of extracted concepts, are incorporated into the prompt. This enables flexible control over the scope and granularity of the extracted semantic concepts.

The extracted outputs correspond to atomic semantic concepts rather than final schema fields. Example concepts extracted from FUEL OIL TANKS forms include:

non-destructive test, berthing compartment, tank top, blast, warped seating surface

These concepts form the semantic foundation for the subsequent clustering and field abstraction stages.

3.1.2 Clustering Tool Invocation. To organize extracted concepts into higher-level semantic groupings, the Field Generator Agent invokes a clustering tool over concept embeddings. For each form type $t \in \mathcal{T}$, the extracted concepts are embedded into a semantic vector space and clustered using K-means across multiple values of K to capture semantic abstractions at different granularities. The agent progressively explores larger values of K until it determines that additional refinements no longer yield meaningful semantic abstractions. The resulting concept groups are then aggregated to form an overcomplete set of candidate semantic groups for downstream field abstraction.

For example, $cluster_1 = \{\text{Ripped, Cracked, Loose}\}$ and $cluster_2 = \{\text{Fall risk, Trip hazard}\}$ may form two semantically coherent clusters representing related operational characteristics. This clustering stage consolidates noisy and overlapping concepts extracted from heterogeneous maintenance narratives.

3.1.3 Field Abstraction. After generating concept groups, the Field Generator Agent abstracts each cluster into a higher-level schema candidate field representing the shared semantic meaning of the concepts within that cluster. The abstraction prompt incorporates both clustered concept instances and representative historical forms, enabling the agent to leverage the operational context in which the concepts appear when generating candidate fields.

For example, $cluster_1 = \{\text{Ripped, Cracked, Loose}\}$ may be abstracted as Condition, while $cluster_2 = \{\text{Fall risk, Trip hazard}\}$ may be abstracted as Hazard.

The union of generated field labels across different concept groups and clustering granularities constitutes the overcomplete candidate field space Φ . Table 1 presents examples of generated candidate fields across multiple ship operational form categories. The resulting candidate field space is subsequently provided to the Structural Optimizer Agent for schema optimization.

3.2 Structural Optimizer Agent

After the Field Generator Agent produces the overcomplete candidate field space Φ , the resulting candidate fields are communicated to the Structural Optimizer Agent, which is responsible for constructing compact, informative, and non-redundant schemas for each form type. The agent formulates schema optimization as a stochastic Markov Decision Process (MDP) and sequentially refines the candidate field space through field selection, field merging, and schema termination decisions. As shown previously in Figure 2, the Field Generator Agent produces not only candidate schema fields but also the field-field redundancy matrix and field-form coverage matrix, all of which are utilized throughout the RL-based schema optimization process described next.

3.2.1 RL Formulation.

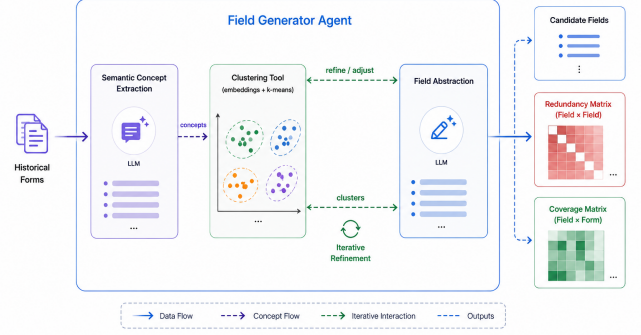


Figure 2: Example snapshot of the Field Generator Agent. The agent extracts semantic concepts and generates candidate schema fields for downstream schema optimization.

State Space. A state is defined as $s = \langle \phi, t \rangle$ where $\phi \in \Phi$ denotes the current candidate schema field and $t \in \mathcal{T}$ denotes the corresponding form type.

Additional states include:

$$\langle \text{START}, t \rangle, \quad \langle \text{END}, t \rangle$$

where $\langle \text{START}, t \rangle$ initializes schema construction for form type t , while $\langle \text{END}, t \rangle$ terminates the current schema generation episode.

Action Space. The action space consists of:

$$\mathcal{A} = \{\text{ADD}, \text{MERGE}, \text{STOP}\}$$

where:

- **ADD** adds the next candidate field to the current schema,
- **MERGE** merges the current field with the next candidate field,
- **STOP** transitions the agent to the terminal $\langle \text{END}, t \rangle$ state.

3.2.2 Environment Transition Modeling. We model the RL environment as a stochastic environment whose transition probabilities are guided by semantic priors extracted from historical operational forms. Figure 3 illustrates a snapshot of the schema construction process.

Given a current state $\langle \phi_k, t \rangle$, an **ADD** action transitions the agent toward another candidate field $\langle \phi_j, t \rangle$, with probability proportional to the occurrence frequency of ϕ_j within historical forms of type t :

$$P_{\text{add}}(\phi_j, t) \propto \text{freq}(\phi_j, t) + \text{noise}_{\text{add}}, \quad \forall \phi_j \in \Phi, \phi_j \neq \phi_k \quad (1)$$

Field frequencies are derived from a binary field-form coverage matrix C generated by the Field Generator Agent:

$$C(\phi_i, f_j) = \begin{cases} 1 & \text{if field } \phi_i \text{ appears in form } f_j \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where field presence is determined through LLM-guided semantic verification of the concepts associated with ϕ_i .

Similarly, a **MERGE** action transitions according to precomputed semantic redundancy statistics:

Form Type	Initial Fields
VOIDS AND COFFERDAMS	Location, Condition, Hazard, Preservation, Accessibility
COMPARTMENTS NOT OTHERWISE COVERED	Condition, Location, Hazard, Mitigation, System, Maintenance, Security, Safety, Bulkhead
STORAGE COMPARTMENTS (incl. ammo, chem, hazmat)	Condition, System Piping, Mitigation, Location, Equipment, Hazard
FUEL OIL TANKS	Leakage, Condition, Mitigation, Location, Leak, Electrical, Electronics
WASTE TANKS	Condition, Location, System, Pumps, Equipment, Piping, System Piping, Tank Condition, System Failure

Table 1: Examples of generated candidate fields across multiple ship operational form categories.

$$P_{merge}(\phi_j) \propto M(\phi_j, \phi_k, t) + noise_{merge}, \quad \forall \phi_j \in \Phi, \phi_j \neq \phi_k \quad (3)$$

where M denotes a field-field semantic redundancy matrix with entries

$$M(\phi_i, \phi_j) \in [0, 1]. \quad (4)$$

Redundancy scores are computed using LLM-assisted semantic comparisons over the concepts associated with each field. Small stochastic noise terms are included to encourage exploration during RL training.

Both field frequency statistics and redundancy scores are pre-computed prior to optimization, separating expensive semantic computations from the online RL process. During each episode, *ADD* and *MERGE* actions continuously update the current schema representation used for reward computation.

3.2.3 Reward Function. The reward function jointly optimizes semantic coverage, schema compactness, and redundancy minimization. Given a transition from state s_t to s_{t+1} under action a , the reward is defined as:

$$R(s_t, s_{t+1}, a) = \alpha \cdot \Delta Coverage - \beta \cdot \Delta Size - \gamma \cdot \Delta Redundancy \quad (5)$$

where

$$\Delta Coverage = Coverage(S_{t+1}) - Coverage(S_t) \quad (6)$$

$$\Delta Size = Size(S_{t+1}) - Size(S_t) \quad (7)$$

$$\Delta Redundancy = Redundancy(S_{t+1}) - Redundancy(S_t) \quad (8)$$

and S_t and S_{t+1} denote the schema before and after the transition, respectively.

Since fields may be merged during optimization, each schema element corresponds to a field group $g_i \subseteq \Phi$ containing one or more candidate fields.

Coverage is computed using the field-form coverage matrix C :

$$Coverage(f_j, g_i) = \frac{1}{|g_i|} \sum_{\phi_k \in g_i} C(\phi_k, f_j) \quad (9)$$

$$Coverage(S) = \sum_{f_j \in \mathcal{H}} \sum_{g_i \in S} Coverage(f_j, g_i) \quad (10)$$

Similarly, redundancy is computed using the field-field semantic redundancy matrix M :

$$Redundancy(g_i, g_j) = \frac{1}{|g_i||g_j|} \sum_{\phi_a \in g_i} \sum_{\phi_b \in g_j} M(\phi_a, \phi_b) \quad (11)$$

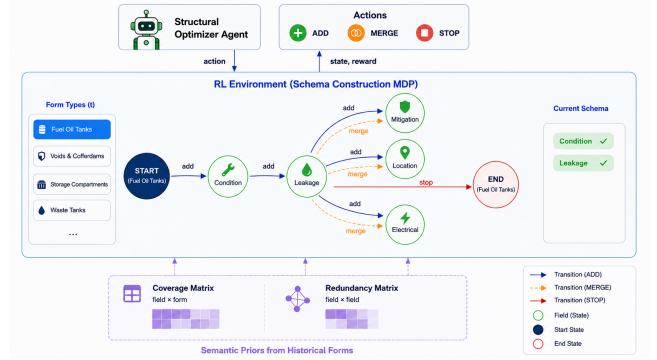


Figure 3: Example snapshot of the Structural Optimizer Agent operating in a stochastic RL environment for schema optimization.

$$Redundancy(S) = \sum_{g_i \in S} \sum_{g_j \in S, j > i} Redundancy(g_i, g_j) \quad (12)$$

And schema compactness is measured as $Size(S) = |S|$.

These quantities are efficiently recomputed after each action and incorporated into the reward signal in Equation 5 during RL optimization.

3.2.4 RL Optimization. We use Temporal Difference Learning (TDL)-based Q-learning for schema optimization. Prior to training, the field-form coverage matrix C and field-field semantic redundancy matrix M are precomputed and used throughout environment transition modeling and reward computation.

The Structural Optimizer Agent learns a Q-table over stochastic schema construction transitions. Q-values are updated according to:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left[r_t + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t) \right] \quad (13)$$

where α is the learning rate, γ is the discount factor, and r_t is the reward from Equation 5.

During inference, the agent starts from $\langle START, t \rangle$ and follows the learned policy until reaching $\langle END, t \rangle$, producing the optimized schema for form type t .

Schema Finalization. After optimization converges, the Structural Optimizer Agent assigns each finalized field group a human-readable label and guide description. This process uses an LLM that receives the merged field groups, representative historical forms,

and pairwise semantic statistics. The resulting schema consists of the finalized field structure together with field-level guidance text supporting human-AI collaboration for form completion during inference.

Schema Population Inference. Once generated, the finalized schema can be applied to unseen operational narratives through a schema population inference process. During inference, an LLM-based Field-Value Extraction Tool leverages the schema fields and their guide descriptions to map unstructured narratives into the generated schema. For each field, the tool extracts a corresponding value when present and returns an associated confidence score.

Human-AI Collaboration. The extracted values and confidence scores can be incorporated into a human-AI collaborative workflow during schema population at inference time. Beyond confidence-aware verification, generated field descriptions can provide field-specific guidance and help generate clarification questions for report authors. Fields with low confidence or missing information may trigger targeted questions that help users provide additional details, improving report completeness and consistency while focusing human effort on uncertain schema components.

3.3 Schema Quality Evaluation

We evaluate schemas generated by **ASMR** using six dataset-driven metrics designed to assess coverage, support, consistency, informativeness, redundancy, and compactness. Table 2 summarizes the corresponding formulations.

Coverage measures how well schema fields represent historical forms, while support captures the probability (P) of observing a schema pattern across the dataset. Consistency measures conditional associations among schema components, and informativeness is estimated using lift-based statistics to quantify non-random relationships between fields. Redundancy quantifies semantic overlap using the precomputed redundancy matrix M , while schema size measures overall schema compactness. Together, these metrics provide a quantitative framework for evaluating the coverage, coherence, compactness, and representational utility of generated schemas.

Metric	Formula
Coverage	$= \frac{1}{ \mathcal{H} S } \sum_{f_j \in \mathcal{H}} \sum_{g_i \in S} \text{Coverage}(f_j, g_i)$
Support	$= P(S)$
Consistency	$= \frac{1}{ S (S - 1)} \sum_{g_i \in S} \sum_{g_j \in S, j \neq i} P(g_j g_i)$
Informativeness	$= \frac{1}{ S (S - 1)} \sum_{g_i \in S} \sum_{g_j \in S, j \neq i} \text{Lift}(g_i, g_j)$
Redundancy	$= \frac{1}{ S (S - 1)} \sum_{g_i \in S} \sum_{g_j \in S, j \neq i} \text{Redundancy}(g_i, g_j)$
Schema Size	$= S $

Table 2: Schema quality evaluation metrics.

4 PRELIMINARY EXPERIMENTAL RESULTS

We conduct a preliminary experimental evaluation of **ASMR** on multiple categories of ship maintenance and operational forms in

Component	Running Time
Field Generator: Concept Extraction	≈ 20 min
Field Generator: Clustering Tool	≈ 5 min
Field Generator: Field Abstraction	≈ 10 min
Field Generator: Coverage & Redundancy Statistics	≈ 35 min
Structural Optimizer: RL Training	≈ 8 sec
Structural Optimizer: Schema Finalization	≈ 2 sec
Inference: Field-Value Extraction Tool	≈ 2 sec

Table 3: Observed running times of major components within the proposed framework.

order to assess the effectiveness of the proposed schema generation and optimization framework.

4.1 Experimental Setup

Experiments were conducted on multiple ship maintenance and operational form categories, including several representative form types presented in Table 1, each containing approximately 500 historical forms.

The Field Generator Agent used GPT-4o Mini for semantic concept extraction and field abstraction, together with embedding-based semantic clustering as an external tool. The Structural Optimizer Agent used TDL-based Q-learning with the stochastic transition model and reward formulation described in Section 3. Hyperparameters, including the learning rate, discount factor, reward weights, and clustering granularity, were selected via grid search.

Algorithms were implemented in Python 3.11 and executed on an HPC cluster consisting of 6 nodes with 2.45 GHz AMD EPYC 7753 processors and 512 GB RAM.

The Field Generator Agent dominates the overall computational cost due to LLM-assisted concept extraction, field abstraction, and construction of the coverage and redundancy statistics required for downstream optimization. In contrast, the Structural Optimizer Agent remains computationally lightweight since these statistics are precomputed offline prior to optimization. Furthermore, inference remains highly efficient, requiring only a few seconds per form.

4.2 Quantitative Results

4.2.1 Metric-Based Evaluation Results. We evaluate the generated schemas using the dataset-driven evaluation metrics introduced in Section 3.3, including coverage, support, consistency, informativeness, redundancy, and schema size. We compare three progressively refined schema representations:

- **Raw Concepts:** atomic semantic concepts directly extracted from historical forms,
- **Candidate Schema:** the overcomplete candidate field space generated by the Field Generator Agent,
- **Optimized Schema (ASMR):** the final schema generated by the Structural Optimizer Agent.

Table 4 summarizes the average metric values across multiple operational form categories.

The results demonstrate a clear progression in schema quality across the three schema representations. Raw semantic concepts

Schema Representation	Coverage	Support	Consistency	Informativeness	Redundancy	Schema Size
Raw Concepts	0.19	< 0.01	0.17	0.94	0.67	42.0
Candidate Schema	0.48	0.08	0.74	1.43	0.39	8.3
Optimized Schema (ASMR)	0.64	0.21	0.69	1.81	0.17	5.4

Table 4: Average quantitative evaluation results across multiple operational form categories.

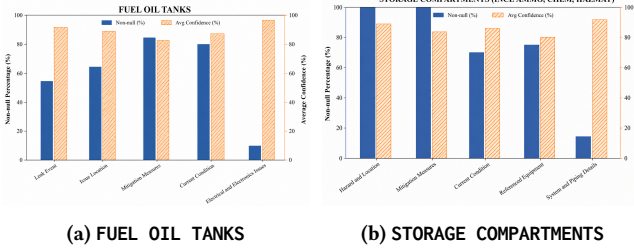


Figure 4: Inference-time field population statistics for finalized schema fields across two operational form categories.

produce highly fragmented and redundant schemas with low statistical support, limited coverage, and weak semantic consistency. After processing by the Field Generator Agent, the candidate schema representation substantially improves coverage, support, consistency, and informativeness while significantly reducing schema redundancy and overall schema size.

The final schemas generated by the Structural Optimizer Agent achieve the strongest overall performance across most evaluation metrics. In particular, the Structural Optimizer Agent significantly improves schema compactness and reduces semantic redundancy while further increasing schema coverage and semantic informativeness. Although the consistency metric slightly decreases after RL optimization, the resulting schemas remain more compact while exhibiting stronger non-random semantic associations. These results demonstrate that the proposed agentic framework can successfully combine semantic field generation and schema optimization to identify compact and semantically meaningful schema representations from heterogeneous operational narratives.

4.2.2 Inference Statistics. To further evaluate the practical usability of the generated schemas during deployment, we analyze inference-time schema population statistics over unseen historical operational forms. During inference, the Field-Value Extraction Tool attempts to populate values for each finalized schema field from incoming operational narratives.

For each form type, the Field-Value Extraction Tool evaluates whether a schema field can be successfully populated from a given form instance together with an associated confidence score for the extracted value. The framework then computes the average non-null presence percentage and average confidence score for each schema field across all forms belonging to the corresponding operational category.

Figure 4 presents an example inference-time analysis for two representative operational form categories, showing both the percentage of non-null extracted values and the average extraction confidence for each finalized schema field.

The results indicate that operationally important schema fields such as mitigation measures, issue locations, and current conditions achieve both high population frequency and strong extraction confidence during inference. In contrast, highly specialized fields such as electrical and electronics issues exhibit lower population frequencies despite maintaining high confidence when present.

These inference statistics provide useful signals for downstream human-AI collaborative operational workflows. In particular, fields with consistently high presence and confidence can be prioritized during AI-assisted form completion, while low-frequency or uncertain fields may trigger targeted guidance, follow-up questions, or user verification to improve report completeness and consistency.

4.3 Qualitative Results

In addition to the quantitative evaluation metrics presented earlier, we further analyze the generated schemas from a qualitative operational perspective in order to better understand how the Structural Optimizer Agent consolidates semantically related fields into compact and interpretable schema representations.

For the Compartments Not Otherwise Covered operational form category, the Structural Optimizer Agent identified that the fields:

Hazard + Safety + Security + System

capture substantially overlapping operational semantics across historical forms. Consequently, the agent merged these fields into a more compact schema representation while preserving the underlying operational meaning. This example further illustrates that the Structural Optimizer Agent can move beyond simple pairwise merging when necessary and progressively consolidate multiple semantically related operational fields into a unified and meaningful schema field.

Similarly, for the Waste Tanks operational form category, the Structural Optimizer Agent identified strong semantic redundancy between fields related to piping and system-level operational descriptions throughout the forms. As a result, fields such as:

Piping + System_Piping

were consolidated into a unified schema component, while redundant operational attributes including overlapping pump and system references were discarded during optimization.

Overall, these qualitative examples demonstrate that the proposed agentic framework not only improves schema quality quantitatively, but also produces compact operational schemas that remain semantically interpretable and operationally meaningful from a human qualitative perspective.

5 RELATED WORK

Our work falls into the intersection of automatic schema generation, agentic systems for data management, and the evaluation of unstructured data processing pipelines.

Automatic Schema Generation. Automatic schema generation is a relevant problem in the context of many systems. Large Language Models and NLP techniques have been used for schema generation within many areas such as relational databases, document-oriented systems, and data warehousing [1, 5, 13]. However, unlike these structured contexts, the free-form nature of ship maintenance report blocks introduces new challenges. Similar AI-driven methods have been developed for schema discovery in JSON documents, knowledge graphs, and ontologies [3, 10, 19], demonstrating both the usefulness of LLMs for schema generation and the contextual improvement schemas can provide for text generation, especially in domain-specific environments [4, 11, 16]. Reinforcement learning [2, 6] has also been proposed for JSON schema generation [9], for fine-tuning rather than schema construction. LLM-driven schema generation thus demonstrates the capability to support structured extraction of maintenance knowledge from historical ship reports.

Agentic Data Systems. The proposed work also relates to agentic unstructured data processing systems [8, 12, 14, 15], as historical ship maintenance reports are largely composed of unstructured text data. With current LLM capabilities, much work has been done in the analysis of unstructured text and optimization of querying and processing of unstructured data. For example, PALIMPZEST [8] proposes a pipeline for the optimization of analytical queries over unstructured data. ELEET [15] focuses on efficient query execution over text and tables using learned multi-modal operators. DocETL [14] proposes agentic optimization and rewriting for complex document processing workflows. DITTO [7] proposes an entity-matching system using pre-trained language models and optimization techniques for domain-specific matching. Our work focuses on a complementary problem, by using unstructured domain-specific data for schema discovery using agents for concept and entity recognition, schema construction, and optimization.

Unstructured Data Pipeline Evaluation. Finally, our work requires a robust evaluation system for determining the quality of constructed schemas. Many schema generation methods for structured data such as in data warehouses [1] employ use of ground-truth or manually designed schemas. In contrast, it is more difficult to attain ground-truth schemas for ship maintenance form types due to lack of standardization. Other methods such as the use of validation agents or NLP techniques are also used for operating on unstructured text data and pipelines [8, 14]. However, use of these methods is highly task-specific and does not directly transfer to our schema evaluation setting. Instead, we construct a quantitative evaluation framework for maintenance report schemas based on core qualities such as coverage, informativeness, compactness, and non-redundancy. Nevertheless, developing more robust methods to address such limitations remains an important direction for future work.

6 CHALLENGES AND OPEN PROBLEMS

Confidence-aware Schema Generation. Large Language Models (LLMs) serve as a key backbone of ASMR, supporting semantic concept extraction, redundancy estimation, and candidate field generation. However, LLMs are inherently probabilistic and may produce inconsistent or semantically ambiguous outputs. An important open problem is the development of confidence-aware schema

generation frameworks that explicitly model uncertainty in AI-generated schema components. One possible direction is to leverage self-consistency prompting and estimate confidence scores over generated concepts and schema fields. Incorporating uncertainty requires jointly reasoning about semantic quality, redundancy, coverage, and confidence, making uncertainty-aware optimization a significant challenge.

Reward Function Design and Quantification. Designing an effective reward function remains a central challenge in ASMR. The current framework balances semantic coverage and redundancy minimization, which often exhibit an inherent tension: increasing schema fields may improve coverage but introduce overlap, while aggressively reducing redundancy may remove operationally important distinctions. Furthermore, redundancy scores are estimated using LLM-based semantic reasoning, which remains subjective and model-dependent. Developing principled formulations for redundancy and other reward dimensions remains an important research problem.

Evaluation. A key challenge faced by ASMR is the absence of reliable ground truth in historical operational datasets. Reports are often noisy, incomplete, and heterogeneous, making semantic interpretation difficult and subjective. While domain experts can assist with validation, manual curation is costly and difficult to scale. Consequently, evaluation is challenging due to both the lack of benchmark datasets and the absence of a definitive “correct” schema for a given form category. Developing objective evaluation criteria and principled validation methodologies therefore remains an important open problem.

Historical Data Limitations. ASMR learns report schemas from historical data, leveraging patterns in past reports to identify information fields for future reporting. However, this approach faces a fundamental bootstrapping challenge: historical data is often incomplete or erroneous due to the very reporting deficiencies that ASMR is intended to address. Consequently, schema-assisted report writing may perpetuate commonly unrecorded information categories. Developing methods to address such limitations, such as data sampling techniques or incorporation of external guidance, remains an important direction for future work.

7 CONCLUSION

This paper presented ASMR, an agentic framework for generating structured schemas for AI-assisted ship operational and maintenance report writing from historical forms. By combining a Field Generator Agent and a Structural Optimizer Agent, the framework automatically identifies compact and semantically meaningful schemas for heterogeneous operational forms. Initial results demonstrate reduced redundancy while preserving semantic coverage and representational quality. Beyond ship reporting, ASMR is applicable to other operational reporting domains. More broadly, this work opens research directions at the intersection of data management, agentic AI, and human-AI collaboration.

ACKNOWLEDGMENTS

This work was sponsored by the Office of Naval Research under Contract No. N0001425C2403

REFERENCES

- [1] Abdelrahman A. Abdelrahman, Abdelrahman A. Elbahrawy, Ahmed R. Sobieh, Alaa E. ElSaid, Ahmed M. Ali, Arwa A. Elsharawy, Yasmine Shaaban, and Yasmine M. Afify. 2025. DataForge: An AI-Driven Data Warehouse Schema Generator. In *2025 Twelfth International Conference on Intelligent Computing and Information Systems (ICICIS)*. 677–684. <https://doi.org/10.1109/ICICIS66182.2025.11313182>
- [2] Andrew G Barto and Richard S Sutton. 2018. *Reinforcement learning: An introduction*.
- [3] Benny Cheung. 2026. Generative Ontology: When Structured Knowledge Learns to Create. *arXiv:2602.05636 [cs.AI]* <https://arxiv.org/abs/2602.05636>
- [4] Yuheng Du, Shereen Oraby, Vittorio Perera, Minmin Shen, Anjali Narayan-Chen, Tagyoung Chung, Anushree Venkatesh, and Dilek Hakkani-Tur. 2020. Schema-Guided Natural Language Generation. In *Proceedings of the 13th International Conference on Natural Language Generation*, Brian Davis, Yvette Graham, John Kelleher, and Yaji Sripada (Eds.). Association for Computational Linguistics, Dublin, Ireland, 283–295. <https://doi.org/10.18653/v1/2020.inlg-1.35>
- [5] Paola Gómez, Rubby Casallas, and Claudia Roncancio. 2020. Automatic Schema Generation for Document-Oriented Systems. In *Database and Expert Systems Applications*, Sven Hartmann, Josef Küng, Gabriele Kotsis, A. Min Tjoa, and Ismail Khalil (Eds.). Springer International Publishing, Cham, 152–163.
- [6] Leslie Pack Kaelbling, Michael L Littman, and Andrew W Moore. 1996. Reinforcement learning: A survey. *Journal of artificial intelligence research* 4 (1996), 237–285.
- [7] Yuliang Li, Jinfeng Li, Yoshihiko Suhara, AnHai Doan, and Wang-Chiew Tan. 2020. Deep entity matching with pre-trained language models. *Proceedings of the VLDB Endowment* 14, 1 (2020), 50–60. <https://doi.org/10.14778/3421424.3421431>
- [8] Chunwei Liu, Matthew Russo, Michael Cafarella, Lei Cao, Peter Baille Chen, Zui Chen, Michael Franklin, Tim Kraska, Samuel Madden, and Gerardo Vitagliano. 2024. A declarative system for optimizing ai workloads. *arXiv preprint arXiv:2405.14696* (2024).
- [9] Yaxi Lu, Haolun Li, Xin Cong, Zhong Zhang, Yesai Wu, Yankai Lin, Zhiyuan Liu, Fangming Liu, and Maosong Sun. 2025. Learning to Generate Structured Output with Schema Reinforcement Learning. *arXiv:2502.18878 [cs.CL]* <https://arxiv.org/abs/2502.18878>
- [10] Felix Neubauer, Benjamin Uekermann, and Jürgen Pleiss. 2025. AI-assisted JSON Schema Creation and Mapping. In *2025 ACM/IEEE 28th International Conference on Model Driven Engineering Languages and Systems Companion (MODELS-C)*. IEEE, 79–83. <https://doi.org/10.1109/models-c68889.2025.00019>
- [11] James Oswald, Kavitha Srinivas, Harsha Kokel, Junkyu Lee, Michael Katz, and Shirin Sohrabi. 2024. Large Language Models as Planning Domain Generators. *arXiv:2405.06650 [cs.CL]* <https://arxiv.org/abs/2405.06650>
- [12] Kiran Kumar Pappula and Guru Pramod Rusum. 2023. Multi-Modal AI for Structured Data Extraction from Documents. *International Journal of Emerging Research in Engineering and Technology* 4, 3 (Oct. 2023), 75–86. <https://doi.org/10.63282/3050-922X.IJERET-V4I3P109>
- [13] Nadia Salem, Khawla Al-Tarawneh, Amjad Hudaib, Hamza Salem, Afaf Tareef, Hadi Salloum, and Manuel Mazzara. 2024. Generating database schema from requirement specification based on natural language processing and large language model. *Computer Research and Modeling* 16 (12 2024), 1703–1713. <https://doi.org/10.20537/2076-7633-2024-16-7-1703-1713>
- [14] Shreya Shankar, Tristan Chambers, Tarak Shah, Aditya G Parameswaran, and Eugene Wu. 2024. Docetl: Agentic query rewriting and evaluation for complex document processing. *arXiv preprint arXiv:2410.12189* (2024).
- [15] Matthias Urban and Carsten Binnig. 2024. Efficient learned query execution over text and tables [technical report]. *arXiv preprint arXiv:2410.22522* (2024).
- [16] Bailin Wang, Zi Wang, Xuezhi Wang, Yuan Cao, Rif A. Saurous, and Yoon Kim. 2023. Grammar Prompting for Domain-Specific Language Generation with Large Language Models. *arXiv:2305.19234 [cs.CL]* <https://arxiv.org/abs/2305.19234>
- [17] Gerald White, Deep Mistry, Kevin Chhoa, Senjuti Basu Roy, Lingyi Zhang, Adam Bienkowski, and Krishna R Pattipati. 2025. A Computational Framework for Estimating Days of Maintenance Delay of Naval Ships.. In *EDBT*. 1014–1022.
- [18] Jingfei Wu, Junyi Han, and Yujin Gao. 2025. TabAgent: A Multi-Agent Table Extraction Framework for Unstructured Documents. In *2025 5th International Symposium on Artificial Intelligence and Big Data (AIBDF)*. IEEE, 600–607.
- [19] Bohui Zhang, Yuan He, Lydia Pintscher, Albert Meroño Peñuela, and Elena Simperl. 2025. Schema Generation for Large Knowledge Graphs Using Large Language Models. *arXiv:2506.04512 [cs.AI]* <https://arxiv.org/abs/2506.04512>