Thesis Proposal: A Normalization-First Framework for Sound, Complete, and Utility-Ready Open Information Extraction

Chandan Prakash* TCS Research, India IIT Kanpur, India ch.pr@tcs.com

Pavan Kumar Chittimalli TCS Research, India pavan.chittimalli@tcs.com

Arnab Bhattacharya IIT Kanpur, India arnabb@cse.iitk.ac.in

Abstract

Open Information Extraction (OIE) has largely focused on extracting relational tuples from text, yet in its current form remains unsuitable for downstream systems due to the absence of standardized, semantically sound representations. This thesis argues that the field has been addressing extraction as a surface-level prediction problem, leading to outputs that are semantically incomplete and logically ambiguous, particularly in the presence of modality, negation, conditionality, quantification, and attribution. We propose a normalization-first framework that reframes OIE as a structured semantic transformation pipeline, where raw text is first converted into a lossless, canonical form of declarative, active-voice, and irreducible sentence units, and extraction is constrained to atomic unary and binary relations augmented with explicit semantic annotations. Within a Probably Approximately Correct (PAC) learning perspective, we formalize soundness, completeness, and usefulness as approximate yet verifiable guarantees over extraction quality, acknowledging the inherent undecidability of full semantic interpretation. This thesis outlines a feasible research program to develop the theoretical foundations, models, and evaluation protocols required to produce system-ready OIE representations, thereby establishing a principled and executable path toward making OIE directly usable for downstream reasoning and machine interpretability.

1 Introduction

Open Information Extraction (OIE) aims to convert unstructured natural language into relational tuples without predefined schemas (Etzioni et al., 2008). It is widely viewed as a general-purpose semantic interface between text and downstream NLP systems. However, despite substantial progress, OIE in its current form is not directly usable for downstream systems due to a lack of standardized, semantically sound representations. Existing approaches largely treat extraction as a surface-level prediction problem, producing tuples that appear plausible but fail to faithfully capture meaning, especially in the presence of modality, negation, quantification, attribution, and conditionality.

OIE emerged to address the limitations of classical Information Extraction, as formalized in MUC-style systems (Sundheim and Chinchor, 1993; Grishman and Sundheim, 1996), which relied on predefined schemas and domain-specific supervision. By removing schema constraints, OIE enabled scalable extraction from open-domain text (Figure 1). This vision was first realized by TextRunner (Etzioni et al., 2008), and subsequently extended through rule-based (Gashteovski et al., 2017; Mishra et al., 2023), neural (Yu et al., 2021; Dong et al., 2023), and generative paradigms (Fan and He, 2023; Zhang et al., 2025). While these approaches improve flexibility and benchmark performance, they do not address a fundamental issue: the absence of a representation that is semantically faithful and directly usable for reasoning.

Current OIE systems implicitly assume that sentences can be decomposed into independent facts. This assumption breaks in realistic settings. For instance, a sentence such as *"The system may automatically approve the loan if the eligibility criteria are met."* is often reduced to $(system, approve, loan)$, discarding modality and conditional dependence. Such outputs are syntactically valid but semantically unsafe, as they conflate possibility with certainty and ignore logical structure. Similar failures arise across benchmarks, where conditionals, negations, and scope are not preserved (Table 1). As a result, improvements in standard evaluation metrics do not correspond to improvements in faithfulness, completeness, or downstream utility (Lamarche and Langlais, 2024).

This problem is further amplified by fragmen-

*External PhD Student at IIT Kanpur

Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 4: Student Research Workshop), pages 1291–1304
July 2-7, 2026

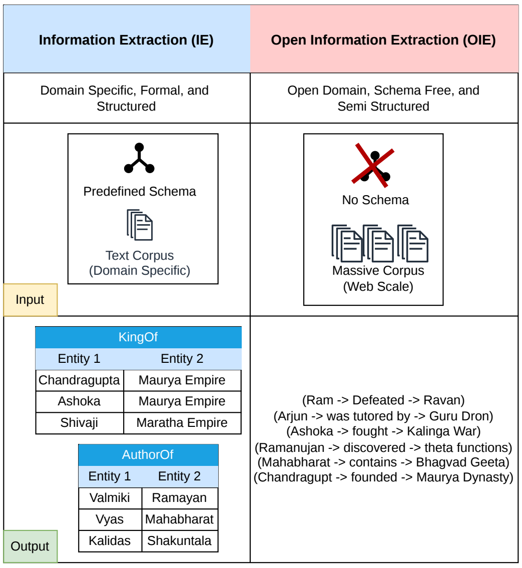

Figure 1: IE vs OIE Comparison

Sentence	*Unless the server is rebooted, the update will not install.*
Extractions	`(the server; is; rebooted), (the update; will not install; )`
Remarks	Conditional dependencies are not represented within the current OIE framework.

Table 1: Representative output from an OIE system demonstrating gaps in soundness, completeness, and downstream usefulness.

tation across representations and evaluation practices. OIE systems vary in extraction formats, granularity, and assumptions about linguistic structure, while benchmarks adopt incompatible annotation schemes. Existing metrics are predominantly string-based and fail to capture semantic correctness or logical consistency. Prior surveys (Niklaus et al., 2018; Zhou et al., 2022; Pei et al., 2023b; Pai et al., 2024) document this diversity, but do not provide a unified formulation grounded in explicit semantic requirements for downstream use. Consequently, the field has largely optimized for extractability rather than usability.

This thesis contends that the central challenge in OIE is a mis-specified problem formulation: it should not be modeled as direct tuple prediction from raw text, but as a structured semantic transformation that yields sound, complete, and downstream-usable representations. Given the ambiguity and partial undecidability of natural language, these properties are defined and enforced within a PAC framework (Haussler and Warmuth, 2018), providing approximate yet verifiable guarantees.

To this end, we propose a normalization-first OIE framework that decouples semantic normalization from relational extraction. The framework comprises four stages (Figure 2). This thesis presents a concrete and feasible research program to operationalize this formulation, with the following primary contributions:

- A formalisation of lossless semantic decomposition (Table 3) with guarantees of preservation and consistency within a PAC framework.
- A standardized semantic annotation schema capturing logic-critical aspects such as modality, scope, negation, and attribution (Table 2).
- The development of benchmark datasets and fine-grained evaluation metrics that measure semantic soundness, completeness, and utility beyond surface matching.
- A system design that integrates probabilistic models with deterministic validation under a PAC framework to ensure reliable extraction.
- Canonicalization and empirical evaluation of OIE representations for downstream reasoning and machine interpretability.
- Extension of the normalization-first framework to multilingual settings to validate its generality across languages.

This work redefines OIE as a problem of producing system-ready semantic representations rather than surface-level extractions. It presents a principled and executable path toward making OIE outputs sound, complete, and directly usable for downstream NLP systems.

2 Related Work

Since *TextRunner* (Etzioni et al., 2008), OIE has evolved through rule-based, neural, and generative paradigms, yet no generation has produced representations that are directly usable by downstream reasoning systems. The core issue is not architectural but representational: the field has treated OIE as a surface-level prediction task, producing tuples that are syntactically plausible but semantically unsound. We survey the literature along four dimensions to establish where and why this gap persists, and to position the proposed normalization-first framework as the necessary corrective.

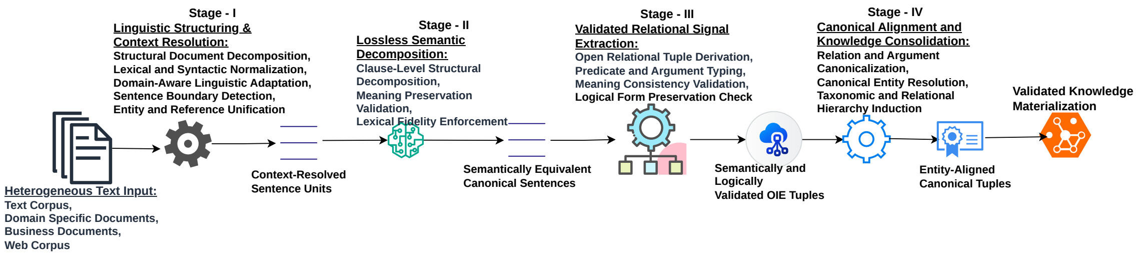

Figure 2: Proposed End-to-End OIE Pipeline for Knowledge Base Population

2.1 OIE System Architectures

Rule-Based Systems: Early systems grounded extraction in syntactic parsing, POS tagging, and semantic role labeling. *WOE* (Wu and Weld, 2010) and *ReVerb* (Fader et al., 2011) improved relation-phrase coherence. *Ollie* (Mausam et al., 2012) partially captured attribution and clausal context. *Kraken* (Akbik and Löser, 2012) introduced N-ary tuples; *ClausIE* (Del Corro and Gemulla, 2013) and *StanfordIE* (Angeli et al., 2015) decomposed sentences into clauses and entailments before extraction. *ReNoun* (Yahya et al., 2014) and *RelNoun* (Pal and Mausam, 2016) extended relational coverage to nominalizations. *NestIE* (Bhutani et al., 2016) encoded hierarchical propositional dependencies via recursive representations. *MinIE* (Gashteovski et al., 2017) was the first to attach structured semantic annotations; polarity, modality, attribution, and quantity; to extracted tuples. *IndIE* (Mishra et al., 2023) extended rule-based extraction to morphologically rich languages. Despite interpretability, these systems are constrained by handcrafted patterns and provide no formal guarantees on semantic preservation.

Neural Tagging Systems. The second generation reframed OIE as sequence labeling. *NeuralOIE* (Cui et al., 2018) applied copy-attention over encoder representations. *OpenIE6* (Kolluru et al., 2020a) combined grid labeling with iterative inference for N-ary extractions. *SpanOIE* (Zhan and Zhao, 2020) improved argument boundary detection at the span level. Later models (Yu et al., 2021; Dong et al., 2023, 2022; Fatahi Bayat et al., 2022) added syntactic constraints to promote compactness. These models improve coverage but cannot represent discontinuous arguments or enforce semantic correctness on outputs.

Neural Generative Systems. Seq2Seq models offer greater expressiveness. *IMOJIE* (Kolluru et al., 2020b) conditioned generation on previously decoded tuples to reduce repetition. *Gen2OIE* (Kolluru et al., 2022) extended this to multilingual settings. More recent models (Pei et al., 2023a; Fan and He, 2023; Chen et al., 2024) leverage pre-trained encoder-decoder architectures. While enabling abstractive and discontinuous extraction, generative systems impose no structural or semantic constraints on outputs, making soundness and completeness unverifiable.

LLMs as OIE Systems. Foundational models (Google; OpenAI; Anthropic Claude; DeepSeek; Alibaba Cloud; xAI; Cohere, 2024) have been applied to OIE in zero and few-shot regimes. *Mario* (Zhang et al., 2025) uses chain-of-thought reasoning to construct binary extractions iteratively. Although LLMs achieve competitive benchmark scores, they hallucinate, are prompt-sensitive, and their outputs have not been evaluated for semantic soundness or logical completeness within the PAC learning framework; the gap most directly motivating this thesis.

2.2 Representation: Structure, Granularity, and Semantic Annotations

Binary tuples dominate due to benchmark compatibility (Stanovsky and Dagan, 2016; Bhardwaj et al., 2019), but compress complex assertions into pairwise relations, causing information loss; N-ary methods (Akbik and Löser, 2012; Del Corro and Gemulla, 2013) improve fidelity yet lack clear criteria for use, while nested representations (Bhutani et al., 2016) remain largely unevaluated. More critically, logic-bearing phenomena such as negation, modality, conditionality, attribution, and quantification are typically discarded, as noted by *MinIE* (Gashteovski et al., 2017) and *Ollie* (Mausam et al., 2012), but inconsistently encoded and absent from benchmarks, reducing conditional tuples (Table 4) to misleading independent assertions. Additionally, sentence normalization is underdeveloped: clause-based approaches *ClausIE* (Del Corro and Gemulla,

2013) and *StanfordIE* (Angeli et al., 2015) lack guarantees of meaning preservation, while simplification methods (Narayan et al., 2017; Ponce et al., 2024; Yao et al., 2024) prioritize readability over semantic fidelity, and *OIE@OIA* (Wang et al., 2022b) introduces intermediate graphs without ensuring lossless decomposition; this thesis instead formalizes lossless normalization within a PAC framework as a prerequisite for reliable extraction.

2.3 Evaluation Benchmarks and Their Limitations

Evaluation has progressed from manual inspection (Mesquita et al., 2013; Xu et al., 2013) to large-scale automated matching, with *OIE2016* (Stanovsky and Dagan, 2016) introducing silver-standard lexical soft matching, and *Wire57* (Lechelle et al., 2019) and *CaRB* (Bhardwaj et al., 2019) advancing toward gold-standard N-ary annotation; subsequent resources such as *AW-OIE* (Stanovsky et al., 2018), *RE-OIE2016* (Zhan and Zhao, 2020), and *LSOIE* (Solawetz and Larson, 2021) support neural evaluation, while specialized benchmarks address Chinese (Sun et al., 2018; Chen et al., 2024), multilingual fact synsets (Gashteovski et al., 2022), pre-trained language model probing (Wang et al., 2022a), cross-domain robustness (Qi et al., 2023), and domain generalization (Yu et al., 2022), with *BenchIE FL* (Lamarche and Langlais, 2024) aligning evaluation with downstream utility via flexible matching; however, three limitations persist: reliance on lexical overlap that ignores semantic equivalence and logical correctness, lack of evaluation for semantic annotations such as polarity, modality, and conditionality, and inconsistent annotation conventions across benchmarks that hinder reliable comparison (Lamarche and Langlais, 2024), so reported gains do not reliably reflect improved system-readiness for reasoning.

2.4 Multilinguality

Multilingual OIE has been explored via rule-based systems for Spanish and Portuguese (Gamallo and Garcia, 2015) and for Hindi, Tamil, Telugu, and Urdu (Mishra et al., 2023); neural tagging for Spanish, Portuguese (Ro et al., 2020; Vasilkovsky et al., 2022), Arabic, and Galician (Kotnis et al., 2022); and generative alignment for cross-lingual transfer (Kolluru et al., 2022). Benchmark resources exist for Chinese (Sun et al., 2018; Gashteovski et al., 2022; Chen et al., 2024), German (Gashteovski et al., 2022), Arabic, Galician, and Hindi. Yet OIE for morphologically diverse languages remains underexplored, and no work has applied normalization-based preprocessing to non-English languages where coreference, implicit reference, and clause boundary detection are substantially harder. We outline multilingual extension as a future direction. Preliminary investigation is required to assess how normalization assumptions transfer across typologically diverse languages.

2.5 Downstream Utility

The ultimate criterion for OIE is downstream utility across tasks such as knowledge base population, question answering, relation discovery, natural language inference, and machine reading comprehension (Mausam, 2016); however, *BenchIE FL* (Lamarche and Langlais, 2024) shows that benchmark scores poorly predict application performance, indicating that optimizing existing metrics does not yield system-ready representations. A key bottleneck is canonicalization: while prior work addresses relation and entity normalization (Christensen et al., 2011), the joint alignment of tuples enriched with polarity, modality, conditionality, and attribution remains unresolved. This thesis introduces a canonicalization and consolidation stage, followed by evaluation on natural language to logical reasoning tasks, providing an executable path toward usable outputs. Existing surveys (Niklaus et al., 2018; Zhou et al., 2022; Pei et al., 2023b; Pai et al., 2024) cover architectural, neural, application, and chronological views, yet overlook the core representational deficit, as successive methods improve benchmarks without resolving semantic incompleteness and logical ambiguity; this thesis instead formulates OIE as a structured semantic transformation pipeline with PAC-grounded guarantees on soundness, completeness, and utility.

3 Proposed Framework and System

We contend that the primary limitation of existing OIE systems lies not in extraction accuracy, but in *representation design*. Specifically, current approaches generate surface-level tuple fragments (Examples in Figure 1 and Table 1) that lack semantic grounding and structural information, as shown in Table 4, making them unsuitable for downstream reasoning or inference tasks.

To address this, we reconceptualize OIE not as a span-labeling or relation-detection problem, but

Table 2: Semantic Annotation Schema for Unary and Binary OIE Tuples

Attribute	Definition and Possible Values
Tuple ID	Unique identifier $(T_1, T_2, \ldots)$ used for referencing tuples in scope and dependency layers.
Type	**Unary**: $P(x)$ representing a property or state. **Binary**: $R(x, y)$ representing a relationship between two arguments.
Predicate	The verb phrase (span or non-span).
Arg 1 (Subject)	Noun phrase (span or non-span) representing the first argument.
Arg 2 (Object)	Noun phrase (span or non-span) representing the second argument, or `NULL` for unary tuples.
Polarity	**Positive** (Assertion), **Negative** (Negation or Denial), **Neutral** (Ambiguous or underspecified).
Modality	**Fact** (Is), **Possibility** (May / Can), **Obligation** (Must / Should), **Necessity** (Will / Essential), **Permitted**, **Prohibited**.
Quantification	**Singular** (Default/One), **Universal** (All / Every), **Existential** (Some / A few), **Cardinal** (Specific number), **None**.
Attribution	Author (Default), an explicitly stated entity acting as the source of the claim, or SYSTEM_INFERENCE
Scope & Dependency	Conjunction / Disjunction: `AND(ID_list)`, `OR(ID_list)` Conditional: `IF(Antecedent_IDs)`, `THEN(Consequent_IDs)`

as a normalization-first problem. In this view, the task involves systematically transforming unconstrained natural language sentences into simplified, normalized sentences (as illustrated in Table 3), followed by their conversion into a semantically faithful and structurally standardized OIE representation (Defined in Table 2), as shown in Table 4.

The proposed framework operationalizes this perspective through a four-stage pipeline (Figure 2), where each stage enforces a verifiable linguistic contract over its output. The pipeline prioritizes text normalization, motivated by the observation that although existing OIE frameworks claim domain independence, empirical studies (Qi et al., 2023; Yu et al., 2022) demonstrate that structural heterogeneity in raw corpora (e.g., business documents versus web-scraped text) adversely affects extraction performance. Accordingly, the normalization stage standardizes sentence structure to produce consistent and well-formed sentence units. The framework is designed to satisfy the key properties of *soundness*, *completeness*, and *usefulness* in the extracted information.

Table 3: Sentence Simplification to Declarative, Active Voice, and Atomic Irreducible Simple or Complex Sentences.

Original Sentence	Simplified Sentences
The system may automatically approve or reject a loan application if the applicant meets all eligibility criteria.	S1: The system may automatically approve a loan application if the applicant meets all eligibility criteria. S2: The system may automatically reject a loan application if the applicant meets all eligibility criteria.
According to the company policy, no employee shall access customer records without authorization.	S3: No employee shall access customer records without authorization.
According to Dr. Smith, the medication must be taken after meals, but it should not be consumed if the patient feels dizzy.	S4: Dr. Smith states that the patient must take the medication after the patient eats meals. S5: Dr. Smith states that the patient should not consume the medication if the patient feels dizzy.

3.1 Formal Definitions

We formalize these properties at the sentence level to define a principled semantic interface between an unstructured input sentence S and its extracted representation T. This formulation is grounded in the central claim that existing OIE systems fail to produce representations that are directly usable for downstream reasoning due to the absence of standardized and semantically faithful structure. Accordingly, we define the requirements for system-ready representations in terms of soundness, completeness, and usefulness, interpreted within the PAC learning framework. The proposed representation is related to Semantic Web formalisms such as RDF (Resource Description Framework) triples and reification, but differs in its explicit modeling of modality, attribution, and conditional scope, which are typically not represented within such frameworks.

Soundness (Semantic Faithfulness): An extraction is sound if every tuple $t \in T$ is supported by the meaning of the source sentence S, i.e., $S \models t$. This ensures that the extracted representation does not introduce unsupported or hallucinated information. Given the inherent ambiguity and variability

Table 4: Illustrative OIE Tuple with Semantic Annotations. Sentence ID refers to Table 3; annotation schema defined in Table 2.

Sent.	Tuple	Predicate	Argument 1	Argument 2	Pol.	Modality	Attribution	Scope & Dependency
S3	T1	access	employee *Quant: All*	customer records *Quant: All*	-ve	Prohibited	ORG(policy)	IF ([T2]) THEN ([T1])
	T2	have authorization	employee *Quant: All*	NA *Quant:*	-ve	Fact	Author	

of natural language, this guarantee is interpreted within the PAC learning framework, where soundness holds with high probability under a distribution over sentences.

Completeness (Information Preservation): An extraction is complete if the combined semantics of all tuples in T preserves the meaning of the original sentence S, i.e., $\bigwedge_{t \in T} t \approx S$. This requires that critical semantic phenomena such as modality, negation, and conditionality are retained in the transformation. As exact equivalence is generally undecidable for natural language, completeness is defined in a PAC sense, where the representation approximately preserves meaning with bounded error.

Usefulness (Reasoning Readiness): An extraction is useful if the representation T is structured and semantically explicit enough to support downstream reasoning tasks without requiring access to the original sentence S. This includes enabling inference over entities, relations, and their associated semantic constraints. Usefulness reflects the ability of T to support correct reasoning outcomes with high probability.

While we adopt PAC terminology to frame soundness and completeness as approximate guarantees, we operationalize these notions through measurable proxies. In practice, entailment is approximated using a combination of natural language inference models and human validation. Meaning preservation is evaluated via reconstruction fidelity and semantic annotation coverage. Specifically, soundness is evaluated via entailment-based verification between source sentences and extracted tuples, using both language models and human validation. Completeness is approximated through reconstruction fidelity and coverage of annotated semantic phenomena such as modality and negation. Rather than deriving formal PAC bounds, this work interprets PAC as a guiding principle for defining probabilistic guarantees over extraction quality under realistic linguistic uncertainty.

These properties collectively redefine OIE as a representation problem rather than an extraction problem. The tuple extraction examples presented under the existing OIE framework in Table 1 are neither sound nor complete, and they are not directly useful for downstream reasoning tasks. In contrast, the examples under the proposed framework, shown in Table 4, are sound, complete, and useful for downstream reasoning tasks. To operationalize this perspective, we propose a normalization-first pipeline (Figure 2) that systematically transforms unstructured text into semantically grounded, structurally standardized representations suitable for machine interpretability and downstream inference.

3.2 Stage I: Linguistic Structuring and Context Resolution

Despite claims of domain independence, existing OIE systems degrade under structural variability in real-world text (Qi et al., 2023; Yu et al., 2022). This reveals a fundamental issue: OIE has focused on extraction without first ensuring that the input is linguistically well-formed for consistent interpretation.

In contrast, our framework treats preprocessing as a semantic structuring step. Given an input text, this stage produces a set of standardized sentence units with resolved coreference and explicit contextual links. This transformation reduces ambiguity and ensures that each unit is self-contained for downstream processing. Soundness at this stage, interpreted within the PAC learning framework, requires that contextual resolution preserves the intended referential meaning with high probability under a distribution of inputs.

Exploration Status: We have explored both dependency-based rule systems and LLM-based approaches for segmentation and context resolution. Our findings indicate that early resolution of coreference and implicit links is necessary for achieving semantically faithful representations. Ongoing work investigates domain-adaptive preprocessing strategies for structurally complex corpora such as legal and technical text.

3.3 Stage II: Lossless Semantic Decomposition

Natural language sentences exhibit significant functional and structural variability. Functionally, they may be Declarative, Interrogative, Imperative, or Exclamatory; structurally, they range from Simple to Compound-Complex, it can be in Active or Passive voice, which current OIE systems attempt to handle implicitly, leading to inconsistent outputs. We instead formalize decomposition as a normalization problem. Given a sentence S, we construct a set of simplified sentences in Canonical Normal Form (CNF), denoted as

$$\mathcal{S}_{cf} = \{s_1, s_2, \ldots, s_n\},$$

where each s_i is declarative, in active voice, and structurally irreducible.

This transformation is governed by two constraints:

1. **Irreducibility:** Each $s_i \in \mathcal{S}_{cf}$ cannot be further decomposed without breaking a logical dependency. For example, conditional constructions are preserved as single units. This ensures that the decomposition aligns with semantic structure rather than surface syntax.

2. **Bounded Lexical Fidelity:** Let $V(\cdot)$ denote the vocabulary of a sentence. We require that

$$1 - \frac{|V(\mathcal{S}_{cf}) \cap V(S)|}{|V(\mathcal{S}_{cf})|} \leq \epsilon, \qquad (1)$$

   ensuring that introduced tokens are limited to syntactic transformations rather than semantic alterations. This constraint enforces semantic faithfulness while allowing necessary normalization operations.

Completeness of the transformation that $\mathcal{S}_{cf}$ approximately preserves the full meaning of S is verified via two complementary checks: bidirectional entailment ($S \models \bigwedge s_i$ and $\bigwedge s_i \models S$) and reconstruction, where S is regenerated from the aggregate content of $\mathcal{S}_{cf}$ and compared to the original.

Sentence	*Albert Einstein died in Princeton in 1955.*
Existing Framework Extractions	`(Albert Einstein; died)`, `(Albert Einstein; died; in Princeton; in 1955 )`
Proposed Framework Extractions	T1:`(Albert Einstein; died)`, T2:`(Albert Einstein; died in; Princeton )`, T3:`(Albert Einstein; died in; 1955 )`, AND(T2,T3)

Table 5: Decomposition of an n-ary event into compositional unary and binary relations with explicit dependency structure under the proposed normalization-first framework.

Both checks are interpreted as PAC guarantees with bounded approximation error.

Exploration Status: We have experimented with fine-tuned small language models for this transformation. To validate semantic preservation ($S \equiv \mathcal{S}_{cf}$), we explore bidirectional entailment and reconstruction-based checks, both interpreted under PAC guarantees.

3.4 Stage III: Validated Relational Signal Extraction

Once sentences are transformed into canonical form, the extraction problem becomes structurally constrained. This enables a shift from heuristic extraction to schema-driven interpretation. We restrict the representation to atomic **unary** (Entity $\rightarrow$ Property) and **binary** (Subject $\rightarrow$ Object) relations (Table 2), augmented with a semantic annotation layer $\mathcal{A}$ capturing polarity, modality, quantification, condition, and attribution.

The restriction to unary and binary relations is a deliberate design choice to enable standardization, interpretability, and tractable evaluation. Although natural language frequently expresses n-ary or event-centric structures, we posit that a large class of such phenomena can be systematically decomposed into sets of interdependent unary and binary relations, augmented with explicit scope and dependency annotations (Table 5). This compositional representation prioritizes semantic transparency and operational consistency over representational compactness. While certain inherently n-ary constructs may require approximation, the central objective of this work is to determine whether semantically faithful and reasoning-ready representations can be achieved within a constrained yet standardized relational framework.

This design reflects the central thesis: OIE should produce semantically explicit and structurally standardized representations rather than loosely defined tuples. Soundness and completeness of the extracted representation are enforced through consistency with the canonical sentences, interpreted within the PAC framework to account for uncertainty in language understanding.

Exploration Status: Our experiments indicate that rule-based extraction over normalized sentences achieves higher structural consistency than end-to-end neural systems, which exhibit schema drift and introduce vocabulary absent from the source, consistent with the limitations documented in recent survey work (Pei et al., 2023b). Current efforts focus on constructing a fine-grained evaluation benchmark that measures annotation accuracy for $\mathcal{A}$, moving beyond the lexical overlap metrics that dominate existing OIE evaluation (Yu et al., 2022).

3.5 Stage IV: Canonical Alignment and Representation Consolidation

The final stage enforces global consistency across extracted representations by aligning semantically equivalent expressions and resolving variation in surface forms. This process ensures that the output representation is stable, comparable across inputs, and suitable for downstream reasoning.

Rather than treating extracted tuples as isolated fragments, this stage organizes them into a coherent semantic representation with consistent relational forms and entity references. Within the PAC learning framework, this alignment is treated as an approximate consistency guarantee, ensuring that semantically equivalent inputs yield comparable representations with high probability.

Exploration Status: We have explored clustering-based approaches (Vashishth et al., 2018) to identify equivalent relational expressions. Ongoing work focuses on integrating these methods with the earlier stages to produce end-to-end consistent representations, bridging the gap between probabilistic language processing and reliable system-ready outputs.

3.6 Benchmark Construction and Evaluation Protocol

A recurring problem in OIE research is that evaluation is decoupled from the properties that matter. Existing benchmarks measure token-level overlap against human-annotated tuples (Yu et al., 2022), rewarding systems that produce fluent-looking output regardless of whether it is sound, complete, or useful for downstream reasoning. Benchmark scores consequently do not correlate with downstream utility (Pei et al., 2023b). A central contribution of this thesis is therefore a benchmark and evaluation protocol designed specifically to measure the properties defined in Section 3.1.

Corpus Design: The benchmark draws sentences from several typologically distinct domains, including *Web*, *News*, *Literature*, *Banking*, *Law*, *Medical*, *Automobile*, and *Transportation*, among others. This diversity is not incidental, it is the primary mechanism for measuring domain independence, which OIE systems claim but rarely demonstrate empirically. The distribution of sentences will reflect real-world data, encompassing both short and sufficiently long sentences, where most existing OIE systems tend to fail. The corpus is designed to exceed the scale of the current largest fine-grained OIE benchmark, BenchIE FL (Lamarche and Langlais, 2024), which contains only 300 sentences, providing broader coverage for statistically meaningful domain-stratified analysis.

Annotation Schema: Each source sentence S is paired with two layers of annotation: (i) a set of normalised sentences $\mathcal{S}_{cf}$ as defined in Section 3.3, and (ii) a set of gold tuples extracted from $\mathcal{S}_{cf}$ with full semantic annotations $\mathcal{A}$ covering polarity, modality, quantification, and attribution as defined in Section 3.4.

The CNF annotation serves two roles: it defines a reference for normalization quality via a bounded lexical fidelity constraint (Equation 1), ensuring high lexical overlap with S so that both normalization-first and direct extraction systems remain comparably evaluable on the same benchmark, and it provides reusable, high-overlap sentence variants that support adjacent tasks such as sentence simplification, textual entailment, and machine translation evaluation.

3.6.1 Evaluation Metrics.

We define a hierarchical scoring scheme that maps directly onto the formal desiderata. For each sentence, partial and full credit is assigned across four dimensions:

Argument Identification Score: measures whether the extracted arguments (subject, object, entity) match the gold arguments, assessed via lexical and semantic overlap.

Relation Phrase Score: measures the accuracy of the extracted predicate against the gold relation phrase.

Complete Relation Score: a joint score over the full tuple $(subject, predicate, object)$, rewarded only when all three components are jointly correct, forming the basis for the soundness requirement.

Semantic Annotation Score: measures the accuracy of the annotation layer $\mathcal{A}$ polarity, modality, quantification, and attribution, against gold labels, operationalising completeness for the logical phenomena that current benchmarks ignore entirely.

An aggregate system score is computed as a weighted combination of four dimensions, with domain-stratified scores reported separately to expose domain generalization gaps hidden by aggregate metrics; this reflects the central premise of the thesis that OIE systems should be evaluated on their ability to produce structurally coherent, semantically correct tuples across domains, rather than optimizing token overlap on narrow benchmarks.

The proposed evaluation metrics serve as empirical proxies for PAC-style guarantees. Specifically, tuple-level soundness is assessed through the combination of complete relation accuracy and semantic annotation correctness, ensuring that extracted tuples are both structurally and semantically faithful to the source. Completeness is approximated via semantic equivalence, sentence reconstruction accuracy, and entailment-based consistency, which together measure the extent to which the full meaning of the source sentence is preserved. Usefulness is reflected through aggregate performance, including evaluation on explicit downstream applications, where task-specific objectives provide a direct measure of the practical utility of the extracted representations.

We emphasize that the proposed framework is not intended as a fully deterministic or complete semantic parser. Instead, it defines a structured approximation pipeline where each stage incrementally improves semantic fidelity under practical constraints. The goal is not perfect normalization, but controlled and measurable improvements over existing extraction paradigms.

Conclusion

This thesis argues that OIE has been misformulated as a surface-level extraction task, resulting in outputs that are not directly usable for downstream systems due to the lack of standardized, semantically sound representations, and proposes a normalization-first framework that redefines OIE as a structured semantic transformation problem. By introducing lossless decomposition into canonical sentence units, constraining extraction to atomic relations enriched with explicit semantic annotations, and enforcing consistency through canonical alignment, the framework ensures that meaning is preserved and made interpretable for reasoning, with guarantees defined and evaluated within a PAC learning framework. It further establishes a concrete and feasible research program to formalize decomposition, design a unified annotation schema, develop benchmarks and metrics aligned with semantic soundness and utility, and build models that integrate learning with validation under PAC constraints. Together, this work provides a principled and executable path to transform OIE from heuristic tuple generation into a system-ready semantic interface, directly addressing the core limitation of current approaches and repositioning the field toward representations that are reliable, complete, and usable for downstream NLP tasks.

Limitations

This work has several limitations that reflect both the complexity of natural language and the scope of the proposed formulation. First, the normalization-first framework depends on the reliability of upstream linguistic processes such as coreference resolution and syntactic parsing, and errors in these components can propagate and affect the semantic fidelity of the final OIE output, even when guarantees are defined within a PAC learning framework. Second, enforcing lossless semantic decomposition and explicit annotation introduces additional computational overhead, creating a trade-off between efficiency and semantic soundness that must be carefully managed within PAC-constrained approximations. Third, the restriction to unary and binary relations, while enabling standardization and interpretability, may limit the direct representation of inherently n-ary or event-centric phenomena, requiring further abstraction or decomposition that may not always be minimal or unique. Fourth, although the proposed semantic annotation schema is designed to capture core aspects such as modality, negation, scope, and attribution, it may not fully cover all linguistic nuances, particularly in informal, noisy, or highly context-dependent text, highlighting the limits of any fixed schema under the inherent ambiguity of language within a PAC setting. Finally, the evaluation of semantic soundness, completeness, and usefulness relies on carefully constructed annotated datasets, which are costly, potentially subjective, and difficult to standardize at scale, making robust and generalizable evaluation an open challenge for the proposed research program.

References

Alan Akbik and Alexander Löser. 2012. KrakeN: N-ary facts in open information extraction. In *Proceedings of the Joint Workshop on Automatic Knowledge Base Construction and Web-scale Knowledge Extraction (AKBC-WEKEX)*, pages 52–56, Montréal, Canada. Association for Computational Linguistics.

Alibaba Cloud. Alibaba cloud model studio.

Gabor Angeli, Melvin Jose Johnson Premkumar, and Christopher D. Manning. 2015. Leveraging linguistic structure for open domain information extraction. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 344–354, Beijing, China. Association for Computational Linguistics.

Anthropic Claude. Anthropic claude.

Sangnie Bhardwaj, Samarth Aggarwal, and Mausam. 2019. CaRB: A crowdsourced benchmark for open IE. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6262–6267, Hong Kong, China. Association for Computational Linguistics.

Nikita Bhutani, H. V. Jagadish, and Dragomir Radev. 2016. Nested propositions in open information extraction. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 55–64, Austin, Texas. Association for Computational Linguistics.

Zhen Chen, Jingping Liu, Deqing Yang, Yanghua Xiao, Huimin Xu, Zongyu Wang, Rui Xie, and Yunsen Xian. 2024. Exploiting duality in open information extraction with predicate prompt. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, WSDM '24, page 125–133, New York, NY, USA. Association for Computing Machinery.

Janara Christensen, Mausam, Stephen Soderland, and Oren Etzioni. 2011. An analysis of open information extraction based on semantic role labeling. In *Proceedings of the Sixth International Conference on Knowledge Capture*, K-CAP '11, page 113–120, New York, NY, USA. Association for Computing Machinery.

Cohere. 2024. Cohere. Accessed: September 13, 2024.

Lei Cui, Furu Wei, and Ming Zhou. 2018. Neural open information extraction. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 407–413, Melbourne, Australia. Association for Computational Linguistics.

DeepSeek. Deepseek models.

Luciano Del Corro and Rainer Gemulla. 2013. Clausie: clause-based open information extraction. In *Proceedings of the 22nd International Conference on World Wide Web*, WWW '13, page 355–366, New York, NY, USA. Association for Computing Machinery.

Kuicai Dong, Aixin Sun, Jung-Jae Kim, and Xiaoli Li. 2022. Syntactic multi-view learning for open information extraction. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 4072–4083, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Kuicai Dong, Aixin Sun, Jung-jae Kim, and Xiaoli Li. 2023. Open information extraction via chunks. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 15390–15404, Singapore. Association for Computational Linguistics.

Oren Etzioni, Michele Banko, Stephen Soderland, and Daniel S. Weld. 2008. Open information extraction from the web. *Commun. ACM*, 51(12):68–74.

Anthony Fader, Stephen Soderland, and Oren Etzioni. 2011. Identifying relations for open information extraction. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 1535–1545, Edinburgh, Scotland, UK. Association for Computational Linguistics.

Zhiyuan Fan and Shizhu He. 2023. Efficient data learning for open information extraction with pre-trained language models. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 13056–13063, Singapore. Association for Computational Linguistics.

Farima Fatahi Bayat, Nikita Bhutani, and H. Jagadish. 2022. CompactIE: Compact facts in open information extraction. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 900–910, Seattle, United States. Association for Computational Linguistics.

Pablo Gamallo and Marcos Garcia. 2015. Multilingual open information extraction. In *Portuguese conference on artificial intelligence*, pages 711–722. Springer.

Kiril Gashteovski, Rainer Gemulla, and Luciano del Corro. 2017. MinIE: Minimizing facts in open information extraction. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2630–2640, Copenhagen, Denmark. Association for Computational Linguistics.

Kiril Gashteovski, Mingying Yu, Bhushan Kotnis, Carolin Lawrence, Mathias Niepert, and Goran Glavaš. 2022. BenchIE: A framework for multi-faceted fact-based open information extraction evaluation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4472–4490, Dublin, Ireland. Association for Computational Linguistics.

Google. Google gemini models.

Ralph Grishman and Beth Sundheim. 1996. Message Understanding Conference- 6: A brief history. In *COLING 1996 Volume 1: The 16th International Conference on Computational Linguistics*.

David Haussler and Manfred Warmuth. 2018. The probably approximately correct (pac) and other learning models. *The Mathematics of Generalization*, pages 17–36.

Keshav Kolluru, Vaibhav Adlakha, Samarth Aggarwal, Mausam, and Soumen Chakrabarti. 2020a. OpenIE6: Iterative Grid Labeling and Coordination Analysis for Open Information Extraction. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3748–3761, Online. Association for Computational Linguistics.

Keshav Kolluru, Samarth Aggarwal, Vipul Rathore, Mausam, and Soumen Chakrabarti. 2020b. IMoJIE: Iterative memory-based joint open information extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5871–5886, Online. Association for Computational Linguistics.

Keshav Kolluru, Muqeeth Mohammed, Shubham Mittal, Soumen Chakrabarti, and Mausam. 2022. Alignment-augmented consistent translation for multilingual open information extraction. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2502–2517, Dublin, Ireland. Association for Computational Linguistics.

Bhushan Kotnis, Kiril Gashteovski, Daniel Rubio, Ammar Shaker, Vanesa Rodriguez-Tembras, Makoto Takamoto, Mathias Niepert, and Carolin Lawrence. 2022. MILIE: Modular & iterative multilingual open information extraction. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6939–6950, Dublin, Ireland. Association for Computational Linguistics.

Fabrice Lamarche and Philippe Langlais. 2024. BenchIE^FL: A manually re-annotated fact-based open information extraction benchmark. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 8372–8394, Bangkok, Thailand. Association for Computational Linguistics.

William Lechelle, Fabrizio Gotti, and Phillippe Langlais. 2019. WiRe57 : A fine-grained benchmark for open information extraction. In *Proceedings of the 13th Linguistic Annotation Workshop*, pages 6–15, Florence, Italy. Association for Computational Linguistics.

Mausam, Michael Schmitz, Stephen Soderland, Robert Bart, and Oren Etzioni. 2012. Open language learning for information extraction. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 523–534, Jeju Island, Korea. Association for Computational Linguistics.

Mausam Mausam. 2016. Open information extraction systems and downstream applications. In *Proceedings of the twenty-fifth international joint conference on artificial intelligence*, pages 4074–4077.

Filipe Mesquita, Jordan Schmidek, and Denilson Barbosa. 2013. Effectiveness and efficiency of open relation extraction. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 447–457, Seattle, Washington, USA. Association for Computational Linguistics.

Ritwik Mishra, Simranjeet Singh, Rajiv Ratn Shah, Ponnurangam Kumaraguru, and Pushpak Bhattacharyya. 2023. IndIE: A multilingual open information extraction tool for Indic languages. In *Findings of the Association for Computational Linguistics: IJCNLP-AACL 2023 (Findings)*, pages 312–326, Nusa Dua, Bali. Association for Computational Linguistics.

Shashi Narayan, Claire Gardent, Shay B. Cohen, and Anastasia Shimorina. 2017. Split and rephrase. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 606–616, Copenhagen, Denmark. Association for Computational Linguistics.

Christina Niklaus, Matthias Cetto, André Freitas, and Siegfried Handschuh. 2018. A survey on open information extraction. *arXiv preprint arXiv:1806.05599*.

OpenAI. Openai models.

Liu Pai, Wenyang Gao, Wenjie Dong, Lin Ai, Ziwei Gong, Songfang Huang, Li Zongsheng, Ehsan Hoque, Julia Hirschberg, and Yue Zhang. 2024. A survey on open information extraction from rule-based model to large language model. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 9586–9608, Miami, Florida, USA. Association for Computational Linguistics.

Harinder Pal and Mausam. 2016. Demonyms and compound relational nouns in nominal open IE. In *Proceedings of the 5th Workshop on Automated Knowledge Base Construction*, pages 35–39, San Diego, CA. Association for Computational Linguistics.

Kevin Pei, Ishan Jindal, and Kevin Chang. 2023a. Abstractive open information extraction. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 6146–6158, Singapore. Association for Computational Linguistics.

Kevin Pei, Ishan Jindal, Kevin Chen-Chuan Chang, ChengXiang Zhai, and Yunyao Li. 2023b. When to use what: An in-depth comparative empirical analysis of OpenIE systems for downstream applications. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 929–949, Toronto, Canada. Association for Computational Linguistics.

David Ponce, Thierry Etchegoyhen, Jesús Calleja, and Harritxu Gete. 2024. Split and rephrase with large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11588–11607, Bangkok, Thailand. Association for Computational Linguistics.

Ji Qi, Chuchun Zhang, Xiaozhi Wang, Kaisheng Zeng, Jifan Yu, Jinxin Liu, Lei Hou, Juanzi Li, and Xu Bin. 2023. Preserving knowledge invariance: Rethinking robustness evaluation of open information extraction. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 5876–5890, Singapore. Association for Computational Linguistics.

Youngbin Ro, Yukyung Lee, and Pilsung Kang. 2020. Multi^2OIE: Multilingual open information extraction based on multi-head attention with BERT. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1107–1117, Online. Association for Computational Linguistics.

Jacob Solawetz and Stefan Larson. 2021. LSOIE: A large-scale dataset for supervised open information extraction. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2595–2600, Online. Association for Computational Linguistics.

Gabriel Stanovsky and Ido Dagan. 2016. Creating a large benchmark for open information extraction. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2300–2305, Austin, Texas. Association for Computational Linguistics.

Gabriel Stanovsky, Julian Michael, Luke Zettlemoyer, and Ido Dagan. 2018. Supervised open information extraction. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 885–895, New Orleans, Louisiana. Association for Computational Linguistics.

Mingming Sun, Xu Li, Xin Wang, Miao Fan, Yue Feng, and Ping Li. 2018. Logician: A unified end-to-end neural approach for open-domain information extraction. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, WSDM '18, page 556–564, New York, NY, USA. Association for Computing Machinery.

Beth M. Sundheim and Nancy A. Chinchor. 1993. Survey of the Message Understanding Conferences. In *Human Language Technology: Proceedings of a Workshop Held at Plainsboro, New Jersey, March 21-24, 1993*.

Shikhar Vashishth, Prince Jain, and Partha Talukdar. 2018. Cesi: Canonicalizing open knowledge bases using embeddings and side information. In *Proceedings of the 2018 World Wide Web Conference*, pages 1317–1327.

Michael Vasilkovsky, Anton Alekseev, Valentin Malykh, Ilya Shenbin, Elena Tutubalina, Dmitriy Salikhov, Mikhail Stepnov, Andrey Chertok, and Sergey Nikolenko. 2022. Detie: Multilingual open information extraction inspired by object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 11412–11420.

Chenguang Wang, Xiao Liu, and Dawn Song. 2022a. IELM: An open information extraction benchmark for pre-trained language models. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 8417–8437, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Xin Wang, Minlong Peng, Mingming Sun, and Ping Li. 2022b. OIE@OIA: an adaptable and efficient open information extraction framework. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6213–6226, Dublin, Ireland. Association for Computational Linguistics.

Fei Wu and Daniel S. Weld. 2010. Open information extraction using Wikipedia. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 118–127, Uppsala, Sweden. Association for Computational Linguistics.

xAI. xai grok models.

Ying Xu, Mi-Young Kim, Kevin Quinn, Randy Goebel, and Denilson Barbosa. 2013. Open information extraction with tree kernels. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 868–877, Atlanta, Georgia. Association for Computational Linguistics.

Mohamed Yahya, Steven Whang, Rahul Gupta, and Alon Halevy. 2014. ReNoun: Fact extraction for nominal attributes. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 325–335, Doha, Qatar. Association for Computational Linguistics.

Peiran Yao, Kostyantyn Guzhva, and Denilson Barbosa. 2024. Semantic graphs for syntactic simplification: A revisit from the age of LLM. In *Proceedings of TextGraphs-17: Graph-based Methods for Natural Language Processing*, pages 105–115, Bangkok, Thailand. Association for Computational Linguistics.

Bowen Yu, Yucheng Wang, Tingwen Liu, Hongsong Zhu, Limin Sun, and Bin Wang. 2021. Maximal clique based non-autoregressive open information extraction. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9696–9706, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Bowen Yu, Zhenyu Zhang, Jingyang Li, Haiyang Yu, Tingwen Liu, Jian Sun, Yongbin Li, and Bin Wang. 2022. Towards generalized open information extraction. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 1439–1453, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Junlang Zhan and Hai Zhao. 2020. Span model for open information extraction on accurate corpus. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 9523–9530.

Donghai Zhang, SHuangtao Yang, Dong Xiaozheng, Wei Song, and Bo Fu. 2025. MARIO-0.5B: A multi-agent lightweight model for real-time open information extraction in low-resource settings. In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 13085–13094, Suzhou, China. Association for Computational Linguistics.

Shaowen Zhou, Bowen Yu, Aixin Sun, Cheng Long, Jingyang Li, Haiyang Yu, Jian Sun, and Yongbin Li. 2022. A survey on neural open information extraction: Current status and future directions. *arXiv preprint arXiv:2205.11725*.

A Appendix

We define an extraction as a structured object

$$e = \langle r, A, \Sigma \rangle,$$

where r denotes the relation expression, $A = \{a_1, \ldots, a_n\}$ represents an ordered or unordered set of arguments, and Σ denotes an optional set of semantic annotations such as polarity, modality, or attribution. Differences across Open Information Extraction (OIE) systems can be viewed as variations in the constraints imposed on the arity n, the permissible forms of r and A, and whether Σ is included.

Most existing OIE systems focus on binary relations and typically omit semantic annotations. While some approaches extend to n-ary representations, the semantic annotation component remains largely underexplored, and there is currently no standardized benchmark for its evaluation. The examples in Table 6 are described in light of both current system capabilities and the proposed framework, which is essential for ensuring the soundness, completeness, and practical utility of extracted information. In Table 6, Row 1 illustrates a simple binary extraction, while Rows 2 and 3 demonstrate discontinuous and overlapping binary extractions. Row 5 illustrates an N-ary extraction, and Row 9 illustrates a nested extraction. Rows 10–14 present extractions with semantic annotations, while Rows 6, 7, and 8 illustrate implicit extractions.

ID	Sentence	Extractions
1	Driver is authorized to drive the car.	(Driver, is authorized to drive, the car)
2	The cook baked and ate the cake.	(the cook, baked, the cake) (the cook, ate, the cake)
3	Bell distributes electronic and building products.	(Bell, distributes, electronic products) (Bell, distributes, building products)
4	Elvis moved to Memphis in 1948.	(Elvis, moved to, Memphis) (Elvis, moved in, 1948)
5	Albert Einstein died in Princeton in 1955.	(Albert Einstein, died) (Albert Einstein, died, in Princeton, in 1955)
6	NIH's director, Francis Collins	(Francis Collins, [is] director [of], NIH)
7	United States President Obama	(Obama, [is] President [of], United States)
8	Seattle historian Feliks	(Feliks, [is] historian [from], Seattle)
9	A senior official in Iraq said the body, which was found by U.S. military police, appeared to have been thrown from a vehicle.	P1: (body, appeared to have been thrown, ∅) P2: (P1, from, vehicle) P3: (A senior official in Iraq, said, P2) P4: (U.S. military police, found, body)
10	No people were hurt in the fire.	(people, were hurt in, fire) *Negative Polarity*
11	All young rabbits drink milk.	(rabbits, drink, milk) *Not all rabbits*
12	Early astronomers believed that earth is flat.	(earth, is, flat) *Possibility*
13	It has been proven that earth is round.	(earth, is, round) *Certainty*
14	If he wins five key states, Romney will be elected President.	(Romney, will be elected, President) *Consequent* (Romney, wins, five key states) *Antecedent*

Table 6: Different Types of OIE Extractions. The colour coding is used, blue for arguments, orange for relation phrases, and red for additional information.